

**WISSENSCHAFTLICHE BEITRÄGE**  
aus Forschung, Lehre und Praxis  
zur Rehabilitation behinderter Kinder und Jugendlicher  
Herausgegeben von K. Schulte und W. Katein

**KURT HELLER**

# **INTELLIGENZMESSUNG**

**Zur Theorie und Praxis der Begabungsdiagnostik in Schule und Sonderpädagogik  
unter Berücksichtigung neuer Forschungsergebnisse**

**NECKAR-VERLAG GMBH VILLINGEN**

## Vorwort

Die ursprüngliche Intention dieses Buches zielte darauf ab, eigene im Kontext sonderpädagogischer Diagnostik gesammelte Forschungsergebnisse der Praxis zugänglich zu machen. Einschlägige Erfahrungen im Rahmen der (Sonderschul-)Lehrerbildung veranlaßten mich jedoch schon in einem sehr frühen Entwicklungsstadium des Manuskriptentwurfs, dieses – mehr oder minder rein informatorische – Konzept zu revidieren und *didaktische* Gesichtspunkte stärker miteinzubeziehen. So bereitet es immer wieder vielen Pädagogikstudenten, die ohne grundlegende Kenntnisse der Testtheorie an einem psychodiagnostischen Praktikum teilnehmen, erhebliche Anstrengungen, sich eine fundierte, d.h. wissenschaftlich vertretbare Diagnostikpraxis anzueignen. Die größten Schwierigkeiten betreffen hierbei Probleme der Standardisierung psychodiagnostischer Untersuchungen, Fragen der Methodenwahl (Testgütekriterien) sowie die Testauswertung und Befundinterpretation, also die eigentliche Urteilsfindung oder Gutachtenerstellung. Daß diese Schwierigkeiten bei der Einführung in die pädagogische Diagnostik gerade im sonderpädagogischen Anwendungsbereich virulent werden, ist sicher kein Zufall; nirgendwo sonst im schulpädagogischen Bereich ist der Lehrer häufiger auf testpsychologische Hilfen mannigfacher Art angewiesen, wobei intelligenzdiagnostische Hilfen zweifellos den Löwenanteil beanspruchen. Verstärkte Schwierigkeiten sind jedoch dort zu beobachten, wo es gilt, neue Forschungsbefunde und Erkenntnisse in die Praxis zu transformieren. Und wie Erfahrungen in anderen Bereichen der erziehungswissenschaftlichen Ausbildung lehren, sind die skizzierten Probleme keineswegs ein sonderpädagogisches Spezifikum. Die Einarbeitungsschwierigkeiten treten mindestens im gleichen oder noch größeren Maß bei Lehramtskandidaten der Grund- und Hauptschule, der Realschule und des Gymnasiums (Philologen) auf. Nicht zuletzt aus diesen Einsichten sowie engen Kontakten zu Lehrerkollegen aller Schulgattungen im Rahmen der Bildungsberatung entwickelte sich allmählich das so erweiterte Konzept eines einführenden und sachlich informierenden Buches über Theorie und Praxis der Begabungsdiagnostik in der Schule und Sonderpädagogik, wobei die Verzahnung beider Bereiche hier mehr als ein positiver Nebeneffekt ist. Ein interdisziplinärer Informationsfluß scheint heute angesichts der zu beobachtenden Wissensprogression innerhalb der Sozial- und Erziehungswissenschaften dringender denn je geboten; in besonderem Maß gilt diese Forderung für den Bereich der Begabungsdiagnostik.

Während im ersten Teil des Werkes theoretische Fragestellungen dominieren, wird im zweiten Teil eine Reihe empirischer Dateninformationen in bezug auf intelligenzdiagnostische Fragenkomplexe im schulisch-pädagogischen Kontext geboten. Dem didaktischen Grundanliegen folgend sollen diese an Hand von Beispielen aus der diagnostischen Praxis dem Leser möglichst anschaulich Verwertungsmöglichkeiten eröffnen, ohne gleichzeitig ihre sachlich-inhaltliche Funktion zu schmälern. Methodische und verfahrenstechnische Erörterungen werden dabei ihre unterstützende Wirkung nicht verfehlen. Somit dürfte unserem doppelten Anliegen entsprochen sein: über relevante Forschungsergebnisse zu informieren und gleichzeitig deren Umsetzung auf die konkrete Anwendungssituation zu erleichtern.

Das Vorhaben wäre jedoch nicht realisierbar gewesen ohne die kooperative Bereitschaft zahlreicher Lehrer und Pädagogen an den verschiedensten schulischen Einrichtungen, ohne das Wohlwollen vieler administrativen Stellen, vorab der Schuldirektoren, sowie die aktive Unterstützung durch Studenten der Erziehungswissenschaft(en), die durch ihre Examensarbeiten einen beträchtlichen Teil der *empirischen* Dateninformationen beisteuerten. Ihnen allen und den zahlreich beteiligten Schülern gebührt hier Dank und Anerkennung.

Weiterhin danke ich meinen wissenschaftlichen Mitarbeiterinnen, Frau Dipl.-Psych. H. Weinläder, Heidelberg, für ihre bereitwillige Hilfe bei den Testerhebungen an Sonderschulen sowie Frau Dipl.-Psych. B. Minsel und A.K. Gaedike, Bonn, für die kritische Durchsicht des Manuskripts und ihre Hilfe bei der Registererstellung.

Schließlich gilt mein Dank allen Kollegen, die mich in irgendeiner Weise unterstützten, vorab Herrn Prof. Dr. K. Schulte, Heidelberg, ohne dessen Anregung und Ermutigung diese Veröffentlichung wohl nicht zustande gekommen und – nach mehrmaliger Verzögerung – vollendet worden wäre. Die Drucklegung und Herausgabe übernahm dankenswerterweise der Neckar-Verlag in Villingen.

Heidelberg – Bonn, im Sommer 1972

K. H.

# Inhaltsverzeichnis

INTELLIGENZ ALS PROBLEM DER BEGABUNGSFORSCHUNG UND TESTPSYCHOLOGIE . . . . .	8
<b>I. Intelligenz und Begabung . . . . .</b>	<b>9</b>
1. Theorien und ihre Aussageproblematik . . . . .	9
a) Vorerörterungen . . . . .	9
b) Intelligenzdefinitionen . . . . .	9
c) Bedeutungsunterschiede von Intelligenz und Begabung . . . . .	10
d) Phänomenologisch orientierte Intelligenz- und Begabungstheorien . . . . .	12
e) Faktorenthorien der Intelligenz . . . . .	14
Das hierarchische Faktorenmodell der Englischen Schule (Spearman, Burt, Vernon) . . . . .	14
Das Faktorenmodell von R. Meili . . . . .	15
Die multiple Faktorentheorie von L.L. Thurstone (Modell der "Primärfähigkeiten") . . . . .	17
Das Intelligenz-Strukturmodell ("Structure of Intellect") J.P. Guilfords . . .	19
Exkurs: Ergebnisse der modernen Kreativitätsforschung . . . . .	21
Die Untersuchungsergebnisse A.O. Jägers zur Faktorenstruktur der Intelligenz . . . . .	24
2. Die Frage nach der sozio-kulturellen Bedingtheit der Intelligenz im Lichte der neueren Begabungsforschung . . . . .	30
a) Das Problem der Invarianz menschlicher Eigenschaften und Fähigkeiten . . .	31
b) Sozio-kulturelle Determinanten intellektuellen Verhaltens (unter besonderer Berücksichtigung der Sprache) . . . . .	35
Aufbau der Leistungsmotivation . . . . .	35
Formen der Kindererziehung . . . . .	36
Schichtspezifische Sprachkodes . . . . .	38
Die Sapir-Whorf-Hypothese . . . . .	43
Die Kognition Hör-Sprachgeschädigter . . . . .	45
3. Pädagogische und bildungsrelevante Folgerungen . . . . .	47
4. Sonderpädagogische Perspektiven . . . . .	51
<b>II. Einführung in die Testtheorie . . . . .</b>	<b>55</b>
1. Historische Aspekte . . . . .	55
2. Grundstruktur, Gegenstandsdefinition und Phasen des testdiagnostischen Prozesses . . . . .	57



3. Psychometrische Prinzipien und Konzepte (intelligenz)diagnostischer Untersuchungsverfahren . . . . .	59
a) Allgemeine Probleme der Standardisierung . . . . .	59
b) Die Aufgabenanalyse (Itemanalyse) . . . . .	61
c) Testgütekriterien als Merkmale der Brauchbarkeit diagnostischer Untersuchungsverfahren . . . . .	63
Objektivität . . . . .	63
Reliabilität und Standardmeßfehler . . . . .	63
Validität . . . . .	72
d) Das Normenproblem (die Testeichung) . . . . .	75

### **III. Schülerbeurteilung und Bildungsberatung . . . . . 81**

1. Ziele und Aufgaben der Intelligenztestanwendung in Schule und Sonderpädagogik . . . . .	81
2. Komplettierung der testdiagnostischen Erhebung (Design der Datenerfassung) .	83
a) Verhaltensbeobachtung . . . . .	84
b) Fragebogen (Erhebungs- und Beurteilungsbogen) . . . . .	86
c) Psychodiagnostisches Gespräch (Exploration und Anamnese) . . . . .	89
3. Das pädagogische Gutachten (Integration der Testbefunde und Daten- informationen – Beratungshinweise) . . . . .	94

## **FORSCHUNGSBEFUNDE ZUR BEGABUNGSDIAGNOSTIK IN DER SCHULE UND SONDERPÄDAGOGIK . . . . . 111**

### **I. Untersuchungsziele . . . . . 112**

### **II. Methodenverzeichnis (Intelligenz-/Leistungstests, [Persönlichkeits-]Interessentests) . . . . . 113**

1. Tests zur Erfassung der Allgemeinbegabung . . . . .	114
a) Individualtests . . . . .	114
Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE) . . . . .	114
Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK) . . . . .	119
Haptic Intelligence Scale for Adult Blind (HISAB) . . . . .	123
Snijders-Oomen Nicht-verbale Intelligenztestreihe (SON) . . . . .	125
b) Gruppentests . . . . .	128
Aufgaben zum Nachdenken (AzN 4+) . . . . .	128
Culture Free or Culture Fair Intelligence Tests (CFT) . . . . .	130
Progressiver Matrizen-Test (PMT) . . . . .	132
2. Differentielle Intelligenztests . . . . .	135
Leistungsprüfsystem (LPS) . . . . .	135
Prüfsystem für Schul- und Bildungsberatung (PSB) . . . . .	139

3. Tests und Verfahren zur Erfassung spezieller Fähigkeits- und Verhaltensmerkmale	140
a) Individualtests bzw. Einzelverfahren	140
Benton-Test	140
Vocational Intelligence Scale for Adult Blind (VISAB)	141
Konzentrations-Verlaufs-Test (KVT)	143
Persönlichkeits-Interessen-Test (PIT)	144
b) Gruppentests bzw. Gruppenverfahren	146
Frankfurter Wortschatztest (WST 5-6 u. WST 7-8)	146
Fremdsprachen-Eignungstest für die Unterstufe (FTU 4-6)	147
Mechanisch-Technischer-Verständnis-Test (MTVT)	149
Bourdon-Test	150
Konzentrations-Leistungs-Test (KLT)	151
Pauli-Test	152
Berufs-Interessen-Test (BIT)	154
<b>III. Problembereiche und Befunde intelligenzdiagnostischer Untersuchungen in der Schule und Sonderpädagogik</b>	157
1. Intelligenzdiagnostische Probleme in der Grundschule und den weiterführenden schulischen Bildungseinrichtungen (Hauptschule, Realschule, Gymnasium)	157
a) Die Interpretation von Testleistungs- bzw. Profildifferenzen	157
b) Probleme und Methoden der Schuleignungsermittlung	161
Exkurs: Eignungskriterien für den Englischunterricht an der Hauptschule (Hauptschule A-Kurs)	182
2. Intelligenzdiagnostische Probleme in der Sonderpädagogik	185
a) Probleme und Ergebnisse der Lernbehindertendiagnostik	185
b) Probleme und Ergebnisse der Hörgeschädigten-/Sprachbehindertendiagnostik	196
Exkurs: Das Legasthenie-Syndrom	205
c) Probleme und Ergebnisse der Sehgeschädigtendiagnostik	206
<b>IV. Desiderata</b>	218
<b>V. Tabellen-Anhang</b>	221
<b>ZUSAMMENFASSUNG</b>	247
Literaturverzeichnis	253
Verzeichnis der Grafiken	267
Verzeichnis der Tabellen im Text	268
Verzeichnis der Tabellen im Anhang	269
Personenregister	271
Sachregister	275

**INTELLIGENZ ALS PROBLEM**

**DER BEGABUNGSFORSCHUNG UND TESTPSYCHOLOGIE**

# I. Intelligenz und Begabung

## 1. Theorien und ihre Aussageproblematik

### a) Vorerörterungen

Gegenstand unserer Erörterungen ist das intelligente Verhalten. Die Untersuchungen und Publikationen zum Thema *Intelligenz* sind fast Legion. Um so mehr überrascht angesichts dieser Tatsache die Behauptung eines renommierten Psychologen, daß es schwerfalle, eine verbindliche und hinreichend befriedigende Aussage darüber zu machen, was man eigentlich unter „Intelligenz“ zu verstehen habe. Und um die Paradoxie auf die Spitze zu treiben möchte ich hinzufügen, daß Jahr für Jahr Intelligenzmessungen in der Schule, in den psychologischen und medizinischen Beratungsstellen oder im Bereich der Sonderpädagogik – millionenfach – durchgeführt werden.

„Böse Zungen könnten sagen: Eine beneidenswerte Situation: Sie wissen nicht, was es ist. Aber sie können es messen“ (Heiß 1960, S. 6).

Damit, so wird jetzt mancher Leser anzunehmen geneigt sein, könnten wir uns – die Gültigkeit des zitierten Ausspruchs einmal unterstellt – ohne weitere begrifflich-theoretische Erörterung dem eigentlichen Anliegen dieser Veröffentlichung zuwenden, nämlich Fragen der *Erfassung* intelligenter Verhaltensweisen von Kindern und Jugendlichen. Wir tun es insonderheit aus zwei Gründen nicht.

Einmal, um despektierlichen Versuchen zu steuern, psychologische (hier intelligenzdiagnostische) Bemühungen schon *vor* ihrer Effizienzkontrolle in Frage zu stellen. (Analog könnte man entsprechende Unterfangen anderer Wissenschaftsdisziplinen, die sich methodologisch in vergleichbarer Situation wie die erfahrungswissenschaftliche bzw. operational vorgehende Verhaltens-Psychologie befinden, z.B. die moderne Physik, in ihrem praktischen Wert anzweifeln, obschon dies im bezeichneten Falle kaum jemand wagen würde.) Zum andern wird (doch) aus gegenstandstheoretischen Gründen bzw. im Hinblick auf die zu fordernde *Eindeutigkeit* wissenschaftlicher Aussagen<sup>1</sup> sowie aus praktischen Erwägungen heraus eine genauere Fixierung dessen, was operational oder intelligenztestdiagnostisch untersucht werden soll, unerlässlich sein. Ohne vorausgehende begriffliche Klärung der Untersuchungsphänomene wären zumindest Verständnisschwierigkeiten zu erwarten, womit bereits ein weiteres Postulat wissenschaftlichen Tuns, die Notwendigkeit der Befund*mitteilung* qua Voraussetzungsbedingung möglicher Ergebniskontrollen, angesprochen wäre. Eine ausführlichere Diskussion dieser (Grundsatz-)Probleme erfolgte andernorts (Heller 1966, vgl. S. 165 ff.).

### b) Intelligenzdefinitionen

Die meisten der zahlreichen Intelligenzdefinitionen und der ihnen zugrunde liegenden Theorien enthalten wenigstens zwei wichtige Bestimmungsstücke: die biologische versus soziale *Anpassungsfähigkeit an neue* (bislang nicht gelernte bzw. erfahrene) *Problem- oder Aufgabensituationen resp. Umweltgegebenheiten* und das Postulat nach *Ökonomie der Verfügungsmittel* (in der Mittel-Zweck-Relation). Auswahlweise zitieren wir die berühmten Definitionen von Stern und Wechsler sowie neben der phänomenologischen Intelligenzbestimmung von Scheler noch eine neuere Definition von Groffmann.

---

<sup>1</sup> Das wissenschaftstheoretische Postulat der Eindeutigkeit betrifft vorab die semantische und syntaktische (logische) Eindeutigkeit der verwendeten Begriffe und Sätze, im weiteren die Optimierung der Approximation von Theorie (verbal-begrifflich, numerisch-quantitativ oder wie immer symbolisiertem Modell) und Empirie (erfahrbarer Wirklichkeit). Das Ausmaß tatsächlicher Konvergenz entscheidet hier letzten Endes über die Gültigkeit der Aussagen, d.h. darüber, ob die Aussagebefunde und somit die (neuen) wissenschaftlichen Erkenntnisse stimmen.

„Intelligenz ist die allgemeine Fähigkeit eines Individuums, sein Denken bewußt auf neue Forderungen einzustellen; sie ist allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens“ (W. Stern 1912, S. 3).

Und der gleiche Autor später:

„Intelligenz ist die personale Fähigkeit, sich unter zweckmäßiger Verfügung über Denkmittel auf neue Forderungen einzustellen“ (W. Stern 1950, S. 424; 1. Aufl. 1935).

„Intelligenz ist die zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen“ (D. Wechsler 1961, S. 13).

„Ein Lebewesen verhält sich ‚intelligent‘, wenn es ohne Probierversuche oder je neu hinzutretende Probierversuche ein sinngemäßes – sei es ‚kluges‘, sei es das Ziel zwar verfehlendes, aber doch merkbar anstrebendes, d.h. ‚törichtes‘ (‚töricht‘ kann nur sein, wer intelligent ist) – Verhalten *neuen*, weder art- noch individualtypischen Situationen gegenüber vollzieht, und zwar *plötzlich* und vor allem *unabhängig von der Anzahl* der vorher gemachten Versuche, eine triebhaft bestimmte Aufgabe zu lösen“ (M. Scheler 1962, S. 32; 1. Aufl. 1928).

„Intelligenz ist die Fähigkeit des Individuums, anschaulich oder abstrakt in sprachlichen, numerischen und raum-zeitlichen Beziehungen zu denken; sie ermöglicht erfolgreiche Bewältigung vieler komplexer und mit Hilfe jeweils besonderer Fähigkeitsgruppen auch ganz spezifischer Situationen und Aufgaben“ (Groffmann 1964, S. 190).

Im Mittelpunkt begrifflicher Klärung steht somit die Auffassung von der Intelligenz als *Leistungsdisposition*, wobei einmal die spontane Einstellung, dann wieder mehr die reaktive Anpassung oder/und der Instrumentalcharakter (Intelligenz als ausgezeichnetes Mittel zur Selbstbehauptung resp. Weltbewältigung) hervorgehoben werden. Ähnliche oder gleiche Merkmale, teilweise mit Akzentverschiebungen, enthalten die meisten gängigen Intelligenzdefinitionen (vgl. Aurin, Busemann, Engelmayer/Strunz, Gottschaldt, Hofstätter, Lersch, Lückert, Mühle, Piaget, Schenk-Danzinger, Thomae, Wellek, Wertheimer u.a.).

### c) Bedeutungsunterschiede von Intelligenz und Begabung

Die Begriffe Intelligenz und Begabung werden in der Literatur überwiegend synonym gebraucht (z.B. Binet, Ebbinghaus, A. Fischer, Gottschaldt, Stern, Wenzl). Wo dies nicht der Fall ist, treten vor allem zwei Bedeutungsakzentuierungen zutage: Intelligenz im Sinne eines (relativ) *unspezifischen* Fähigkeitskomplexes und Begabung als auf *bestimmte Tätigkeitsfelder* gerichtete Dispositionen (z.B. „musikalische Begabung“, „mathematische Begabung“, „Sprachbegabung“, „Schachbegabung“) <sup>2</sup> versus Intelligenz als der *engere* und Begabung als der *weitere* Begriff (so bei Aurin, Busemann, H. Roth u.a.).

Die Bedeutungsvarianten des Begriffspaars Intelligenz-Begabung unter ersterem Differenzierungsaspekt mögen durch folgende Zitate gestützt werden.

„Die *Intelligenz* ist ein Instrument der Selbstbehauptung und Lebensbewältigung, und zwar, da sie sich den verschiedenen Aufgaben gegenüber in unterschiedlichen Lebenslagen bewährt, *ohne besondere Ausrichtung*“ (Mühle 1969, S. 72; kursiv v. Verf.).

„*Begabung* wird dagegen als eine *Leistungsbereitschaft* angesehen, die in der Hinordnung auf ein bestimmtes Betätigungsfeld selbst erst strukturiert worden ist und ihrerseits wieder durch die *Ausformung* und *Spezifizierung* der Antriebe, Interessen und Motive die Persönlichkeit verändert“ (Mühle 1969, S. 75).

Ähnlich definierte G. Revesz (1952):

„*Begabung*: eine angeborene, durch Übung zu entfaltende Fähigkeit, die in einem begrenz-

---

<sup>2</sup>Kroh (1939, S. 642 f.) spricht in diesem Zusammenhang von *Sonderbegabung*, die vorab in der *Begabungsrichtung* und nur relativ zu anderen Fähigkeitsbereichen auch in der *Begabungshöhe* von der *allgemeinen Begabung* unterschieden wird. Der (ältere) deutsche Sprachgebrauch kennt hierfür noch die griechisch-lateinische Entlehnung „Talent“, eine im Angelsächsischen völlig unspezifisch (häufig) verwendete Vokabel für *Begabung*.

ten Gebiet der menschlichen Tätigkeit den Durchschnitt weit übertreffende Leistungen hervorzubringen vermag" (Zit. nach Mühle, a.a.O.)<sup>3</sup>

Nach dieser inhaltlichen Ausdifferenzierung stünde der Begabungsbegriff dem modernen psychologischen Begriff der *Eignung* nahe, d.h. Begabung im aufgewiesenen Inhalt wäre hier als „Begabung für“ (Mathematik, Musik, Sprachen usw. oder auch eine bestimmte schulische Bildungsform = Schuleignung bzw. einen bestimmten Beruf = Berufseignung) aufzufassen. In Übereinstimmung dazu verwendeten wir die Termini Hauptschuleignung, Realschuleignung, Gymnasialeignung im Sinne operational bestimmbarer, d.h. wahrscheinlicher Begabung für die betr. Schulformen (Heller 1969b, 1970a, 1970b).

Was endlich die zweite Bedeutungsvariation von Begabung – Intelligenz angeht, so möge stellvertretend für ähnliche Konzepte die Einteilung von Mierke in Grund- und Sekundärintelligenzen erwähnt werden. Mierke (1963) unterscheidet zwischen *Kern- oder Grundintelligenz*, die mehr anlagebedingt gedacht wird, und den *Hilfs- oder Stützfunktionen* der Intelligenz, die in stärkerem Maße durch das soziokulturelle Milieu (Schule, Familie u.ä.) geprägt zu sein scheinen. Gedächtnis, Vorstellungsbesitz (Phantasie), Interessen und Motive (besonders Lern- und Leistungsmotivation), Aufmerksamkeit und Konzentration sind solche Stützfunktionen intellektueller Leistungsfähigkeit.

Zweifellos treffen die dargestellten Begriffsinhalte resp. Beschreibungsmodi relevante Aspekte intelligenten oder begabten Verhaltens. Ob damit auch nur näherungsweise die mit „Intelligenz“ und „Begabung“ umschriebenen Phänomene in ihrer Gänze getroffen sind, bleibt mehr als zweifelhaft. Jedenfalls tendieren die modernen Meinungen der Fachleute eher zu einer inhaltlichen Ausweitung als zu einer Einengung der verwendeten Termini, dies gilt besonders für den Begabungsbegriff, wodurch natürlich die Gefahr der Begriffsverwirrung wächst. Da wir uns nun aber – den Regeln der Pragmatik wie der Ökonomie gleichermaßen verpflichtet – einmal terminologisch festlegen müssen, d.h. einerseits eine vernünftige Reduzierung simultan besetzter Begriffsinhalte dringend geboten erscheint, andererseits (um nicht noch größere Verwirrung zu stiften) konventionelle Begriffsfixierungen weitgehend zu berücksichtigen sind, schlugen wir jüngstens vor, die sog. Kern- oder Grundintelligenz Mierkescher Prägung als „Begabung im engeren Sinne“ oder – um Verwechslungen vorzubeugen – als *Intelligenz* zu bezeichnen. Dieser eingeeengte Begabungsbegriff (= Intelligenz) meint vor allem die von der Anlage her (freilich keineswegs ausschließlich, wie noch zu zeigen sein wird) bestimmten Intelligenzfunktionen, z.B. Denken in seinen mannigfachen Formen (sprachlich, numerisch, figural-räumlich bzw. technisch usw.), geistige Kombinationsfähigkeit, logisches Schließen, Beziehungsstiften u.ä.<sup>4</sup>

---

<sup>3</sup>Auf das Erbe-Umwelt-Problem werden wir noch zu sprechen kommen. Siehe auch Roth et al. (1972, S. 131 ff.).

<sup>4</sup>Die enge Beziehung von Denken und Intelligenz wurde in der psychologischen Forschung schon früh erkannt, wenn auch die beiden Forschungsrichtungen über weite Strecken nebeneinander herliefen. Während die *Denkpsychologie* vorab der Prozeßcharakter intelligenter Verhaltens interessierte, z.B. Problemlösungsverhalten, Einfallerleben u.ä. (vgl. Süllwold 1954, 1959; Graumann 1955, 1969), beschäftigte sich die traditionelle *Intelligenzforschung* vorwiegend mit den Ergebnissen des Denkens, den Denkprodukten oder Denk- (Intelligenz-) Leistungen. Dynamik und Inhalte des Denkablaufs fanden auch in früheren Arbeiten zur Gehörlosenpsychologie bemerkenswerte Beachtung (vgl. Frohn 1926, 1927, 1929, 1950 ff.; Höfler 1927, 1952; Affolter 1953; Andre & Gebauer 1955; Kerstan 1957; Deller 1963; weitere Literaturhinweise finden sich bei E. Kern 1958). In unseren Ausführungen hingegen steht die Leistungsthematik (intelligenten Verhaltens) im Mittelpunkt, worauf schon der Haupttitel dieses Buches hinweist. Doch soll die Begriffsdifferenzierung hier – von Intelligenz und Begabung einerseits sowie von Denken und Intelligenz andererseits – nicht auf die Spitze getrieben werden, zumal beide Begriffspaare vielfach in der Literatur (zumal erziehungswissenschaftlicher Provenienz) jeweils synonym gebraucht werden. Im Grunde repräsentieren beide Forschungsansätze (der Intelligenz- und der Denkpsychologie) zwar verschiedene, jedoch eng aufeinander bezogene Aspekte desselben Tatbestandes, nämlich einmal den Verlauf, das andere Mal die Ergebnisse intelligenten versus begabten Leistungsverhaltens.

Demgegenüber könnte man die „Begabung im weiteren Sinne“ oder – vereinfacht – *Begabung* (ohne Spezifikation) artikulieren, um die gesamte Faktorenstruktur der Grundintelligenz (Denkfähigkeiten) einschließlich ihrer Hilfs- und Stützfunktionen terminologisch zu belegen. Dabei fänden wir uns in Übereinstimmung zu den Ergebnissen der heutigen Begabungsforschung, die neben der Anlageseite die Bedeutung sozio-kultureller Determinanten intelligenten Leistungsverhaltens, wie wir noch ausführlich in den nächsten Abschnitten zu erörtern haben (siehe Kap. I, 2 und 3), hervorhebt. So definiert etwa H. Roth (1969) sinngemäß *Begabung* als das Insgesamt personaler und sozial-kultureller Lern- und Leistungsvoraussetzungen, was unseren Begriffsinhalt von „Begabung“ (i.w.S.) freilich eher noch erweitert als einengt. Unter Berücksichtigung der in der Literatur gängigen Terminologie könnte demnach die aufgewiesenen Begriffsdifferenzierungen von Intelligenz und Begabung (vorerst) nur akzentuierend verwendet werden.

#### d) Phänomenologisch orientierte Intelligenz- und Begabungstheorien

Die rein philosophischen, d.h. mehr oder weniger ausschließlich spekulativen Theorien der Intelligenz können wir ohne nennenswerte Nachteile für den Fortgang unserer Untersuchung übergehen. Von den neueren, phänomenologisch orientierten Intelligenztheorien wären vor allem die Anschauungen von Wenzl (1934), Revesz (1952) und Busemann (1955) zu nennen, die wir in unserem Buch über Bildungsreserven würdigten (Heller 1970a, vgl. noch 1969b). Es genügt deshalb, wenn wir hier kurz auf die Begabungstheorie von A. Wenzl – eine der differenziertesten Analysen dieses Genre – eingehen und gelegentliche Ergänzungen aus gestaltpsychologischer Sicht, z.B. sensu Gottschaldt, anbringen. Da nach unserer Erfahrung Kenntnisse über die zuletzt angesprochenen Theorien gerade in Pädagogenkreisen relativ verbreitet anzutreffen, demgegenüber jedoch größere Informationslücken bezüglich der Faktorentheorien der Intelligenz zu erwarten sind, werden wir die zur Verfügung stehenden Seitenkapazitäten mehr auf die operationistischen Ansätze verwenden, was um so mehr gerechtfertigt erscheint, als gerade von diesen fruchtbare Impulse für den empirisch-praktischen Teil der Intelligenzmessung erhofft werden können.

Wenzl (1957) – der ja, wie wir uns erinnern, keinen begrifflichen Unterschied zwischen Begabung und Intelligenz kennt – pointiert in seiner Begabungstheorie den Strukturcharakter bzw. das hierarchische Gefüge der Intelligenz. Er unterscheidet zwischen *Begabungskapazität* (mit ihrer Breiten-, Höhen- und Tiefendimension), *Intelligenztemperament* (Ansprechbarkeit, Spontaneität, Denktempo, Nachhaltigkeit) und *Gedächtnis*. Ähnlich wie vorher Stern (siehe S. 9f.) und später Gottschaldt verweist Wenzl die Intelligenz in das gesamt-personale Gefüge, wobei besonders die *Interessen* bei der Aktualisierung von Begabungs- bzw. Intelligenzleistungen eine entscheidende Rolle spielen.

Gottschaldt (1939, 1969), der sich in seiner theoretischen Analyse sehr eng an das von Wenzl konzipierte Modell anlehnt<sup>5</sup>, kommt in bezug auf eigene umfangreiche Zwillingsuntersuchungen zu dem Ergebnis, daß begabtes Verhalten „ein sehr komplexes, in vielen Schichten der Person begründetes Geschehen“ sei, dessen erbliche Verwurzelung „polymer bedingt“ zu sein scheint (1939, S. 464 u. 514). Danach sind sowohl die Hauptfunktionen der Intelligenz (Begabungskapazität, abstrahierende Denkform, intuitives Denkverhalten, Gedächtnis und Lernen – vgl. Fußnote<sup>5</sup>) als auch, und zwar vermutlich in gesteigertem Maße, die endothyment Grundlagen, wie Aktivitätsgrade (Antriebsstärke u.ä.), Anteilnahme, Aufmerksamkeit, Spontaneität, Denktempo sowie Gefühls- und Stimmungsgrundlagen

---

<sup>5</sup>Gottschaldts Analysemodell enthält 4 Hauptfunktionen, nämlich die *Begabungskapazität* mit ihrer Breiten-, Höhen- und Tiefendimension, womit sowohl die Weite des Wahrnehmungsfeldes als auch der Modus des Wahrnehmungsverhaltens intendiert sind („Merkwelt“ sensu Uexküll), *abstrahierende Denkform* und *intuitives Denkverhalten* („Wesenschau“), die zusammen mehr das statische Gerüst der intellektuellen Ausstattung darstellen, sowie die mehr dynamisch interpretierten Anpassungsfunktionen der Begabung: *Gedächtnis und Lernen*. Diese vier Funktionen sind nach Gottschaldt (1939, S. 473 ff.) in hohem Maße von der Erbanlage bestimmt.

erblich fundiert. Somit spricht eine gewisse Wahrscheinlichkeit dafür, daß die „unteren“ Persönlichkeitsschichten stärker von der Erbanlage abhängig sind als die oberen, noetischen („noetischer Oberbau“ sensu Lersch), die freilich immer noch 2- bis 3fach stärkere Anlagebedingtheit der intellektuellen Funktionen im Verhältnis zur Milieuwirkung erkennen lassen (1939, bes. S. 457 ff. und 512 ff.). In seinen jüngsten Veröffentlichungen, so 1969, hat Gottschaldt zwar – einem gewissen Zeittrend folgend, aber z.T. mit einleuchtenden Argumenten – seine früheren Ergebnisinterpretationen in Richtung erbpsychologischer Abhängigkeit der Begabungsfunktionen erheblich abgeschwächt; trotzdem ist eine Reihe von Beobachtungsphänomenen aus seinen über 30 Jahre (!) andauernden Längsschnittuntersuchungen auch mit Hilfe der modernen Milieutheorie nicht restfrei aufzuklären, so daß man Intelligenzleistungen nach wie vor in gewisser Abhängigkeit von interindividuellen Anlagendifferenzen interpretieren kann. Als Resümee zum Problem Erbe-Umwelt von Begabung bzw. Intelligenz steht heute (nur) gesichert fest, daß Anlage (Erbe) *und* Umwelt (Milieu) die Begabungsleistungen bedingen – eine ähnliche Auffassung kam bereits in der Sternschen Konvergenztheorie zum Ausdruck! –, wohingegen über das Verhältnis ihrer jeweiligen Anteile (bei der Konstituierung von Begabung bzw. Intelligenz) die Meinungen sehr breit streuen, wengleich man in jüngster Zeit die Rolle des soziokulturellen Milieu stärker als früher beachtet. Die Diskussion über diese Fragen hat keineswegs nur akademischen Wert, ihre Konsequenzen berühren fundamentale bildungspolitische Entscheidungen und stoßen hart an schulpädagogische Realitäten, wie in den folgenden Abschnitten noch darzulegen ist. Daß man heute bereits von der (schulischen versus familialen) Möglichkeit des „Begabens“ spricht, beleuchtet schlagartig eine solche Wende in der Einstellung, hoffentlich auch im Einstellungsverhalten der Eltern und Erzieher.

Im Zentrum der begrifflich-deskriptiven Analysen stand die Frage nach dem „Wesen“ der Intelligenz, womit sowohl nach den inneren (strukturellen) Ordnungszügen als auch nach dem Modus gesamtpersonaler Verankerung der Intelligenz, richtiger: der Einordnung beobachteter und beschriebener (begrifflich fixierter) Leistungsphänomene qua Deduktionsbasis der erschlossenen Intelligenz im Person-Umwelt-Bezugssystem, gefragt werden sollte. Daß bislang kein einheitliches Bild über das „Wesen“ der Intelligenz, d.h. in psychologischer Terminologie über Formen und Hintergründe menschlich-intelligenten Verhaltens, zustandekam, liegt vermutlich weniger oder höchstens indirekt an der Begabungsforschung resp. ihren Methodenansätzen als vielmehr an der zu postulierenden Komplexität der beobachteten und beschriebenen „Intelligenz“-Phänomene selber. So können wir jetzt schon festhalten, daß es *die* Begabung oder *die* Intelligenz nicht gibt. Berechtigter wäre die Annahme einer geringeren oder größeren Vielzahl von Intelligenzen, d.h. intellektueller Fähigkeiten oder (ideologiefreier, weil nicht ontologisierend) *Faktoren*, die unser intelligentes Verhalten bewirken. Um dieses Konstrukt „Intelligenz“ in seinem Bedingungs- und Funktionsgefüge präzise zu erfassen, bedarf es – das zeigt die Erfahrung – weiterer und andersartiger methodischer Bemühungen als der bisher referierten; die Erfolge der gerade in den letzten Dezennien durchgeführten operationalen Ansätze weisen auf entsprechende Möglichkeiten hin. Diese Feststellung will die Verdienste spekulativer bzw. phänomenologisch-deskriptiver Phänomenaufklärung gerade in bezug auf das intellektuelle Verhalten keineswegs schmälern, sie bekräftigt jedoch einmal mehr die Notwendigkeit der Konvergenz beider Forschungsrichtungen: des deskriptiven resp. theoretisch erhellenden *und* des erfahrungswissenschaftlichen bzw. operationalen Ansatzes, wie wir schon eingangs (vgl. bes. Fußnote 1) andeuteten.



## e) Faktorentheorien der Intelligenz

Der bisher fruchtbarste operationale Forschungsansatz zur Aufdeckung des Wirk-Funktions-Systems „Intelligenz“ (das für den strengen Operationalisten zunächst nicht mehr als ein gedankliches Konstrukt darstellt) war und ist zweifellos die *Faktorenanalyse*<sup>6</sup>. Diese Methode repräsentiert einen multivariaten Versuchsansatz, der verfahrenstechnisch auf sehr komplizierten Korrelationsberechnungen basiert.

„Bei der Ordnung einer Mannigfaltigkeit von Phänomenen geht die Faktorenanalyse von deren Wechselbeziehungen (Korrelationen) aus. Ihr Ziel ist die Ermittlung der Anzahl und die Charakterisierung der unabhängigen Dimensionen (Faktoren), auf die sich ein solches System von Wechselbeziehungen zurückführen läßt. Die Anzahl dieser Faktoren ist wesentlich geringer als die der analysierten Phänomene und gestattet so deren relativ einfache Ordnung“ (Jäger 1967, S. 7).

Die Fortschritte der modernen Testdiagnostik – und damit die modernen Möglichkeiten der Intelligenzmessung – beruhen zu einem erheblichen Teil auf faktorenanalytischen Ergebnissen; die Anfänge ihrer ersten Versuche reichen bis zur Jahrhundertwende zurück. Die vier oder fünf (auch im Hinblick auf unsere empirischen Intelligenzuntersuchungen unten) bedeutsamsten Faktorentheorien seien nun im folgenden kurz beschrieben.

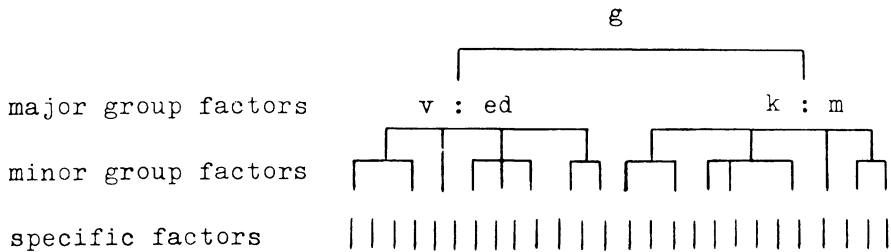
*Das hierarchische Faktorenmodell der Englischen Schule (Spearman, Burt, Vernon u.a.)*

Die Faktorentheorie der Englischen Schule, als deren prominenteste Vertreter Spearman (1904, 1927), Burt (1949) und Vernon (1961) – weitere Literaturhinweise finden sich bei Jäger (1967) – zu nennen sind, geht in ihrer heutigen Fassung über die ursprüngliche *Zwei-Faktoren-Theorie* Spearmans hinaus und versucht (etwa in der Vernonschen Formulierung von 1961), wesentliche Auffassungen multifaktorieller Konzepte (z.B. der „Oligarchic Doctrine“ von Thurstone oder der „Anarchic Doctrine“ von Thorndike) mit dem Spearmanschen Modellansatz zu vereinigen. Spearmans Faktorenmodell der Intelligenz enthielt zunächst einen allgemeinen (general) Faktor und eine unbestimmte Anzahl spezifischer (special) Faktoren. Während dabei dem g-Faktor eine Beteiligung an allen Intelligenzleistungen zugesprochen und in ihm so etwas wie eine zentrale mentale Energie vermutet wird, repräsentieren die s-Faktoren die jeweilige Besonderheit spezieller Leistungsformen. Die Tatsache nicht zu übersehender Überlappungsbereiche dieser s-Faktoren (d.h. ihrer gemeinsamen Varianzanteile *nach* Abzug der auf das Konto von „g“ gehenden Varianzen) bewog jedoch später Burt (1949 ff.) zu einer Revision der ursprünglichen Zweifaktoren-Theorie, indem er die Realität von *Gruppenfaktoren* prinzipiell anerkannte und im modifizierten Modell entsprechend berücksichtigte. Zumindest mußten nach den übereinstimmenden Ergebnissen der Faktorenanalysen in den 40er und 50er Jahren weitere (g-)Faktoren angenommen werden, die – neben dem traditionellen Generalfaktor – für die übrigen, sonst nicht aufklärbaren gemeinsamen Varianzanteile der Spezialfaktoren verantwortlich zeichnen. Um jedoch an der Ausgangshypothese eines einzigen g-Faktors festhalten und andererseits weitere Komplexitätsfaktoren mit in den Ansatz hereinnehmen zu können, postulierte man jetzt verschiedene Generalitätsebenen, was zum *hierarchischen Intelligenzmodell* führte. Dieses ist nach Vernon (1961, S. 23) vorläufig <sup>7</sup> folgendermaßen zu denken. Der „general intelligence“ (g-Faktor) sind die sog. „major group factors“ und diesen die sog. „minor group factors“ als übergreifende Einheiten der „specific factors“ (s-Faktoren) untergeordnet. In der Modellzeichnung sähe dies etwa so aus (vgl. auch Jäger 1967, S. 144):

---

<sup>6</sup>Siehe dazu Hofstätter/Wendt (1966), Überla (1968); vgl. auch Jäger (1967), Pawlik (1968), Guilford & Hoepfner (1971). Nach Abschluß dieses Buchmanuskripts erschienen noch Roth et al. (1972) sowie Wewetzer (1972), worauf wir gern empfehlend hinweisen.

<sup>7</sup>Vernon (1961, S. 23) betont, daß es sich um ein erstes Näherungsmodell („first approximation to mental structure“) handelt, das noch zahlreicher Verbesserungen bedarf.



Die beiden Hauptfaktoren (major group factors) tragen die Bezeichnung v : ed bzw. k : m<sup>8</sup>. Der v:ed-Gruppeneinheit ordnet Vernon die Unterfaktoren verbal und numerical zu, der k:m-Einheit die Unterfaktoren mechanical information, spatial und manual. Zwischen den einzelnen Ebenen dürfen freilich keine scharfen Trennungsstriche gezogen werden (wozu vielleicht die Schemazeichnung verleiten könnte), vielmehr sind – nach Meinung von Burt und Vernon – fließende Übergänge oder gar Überschneidungen anzunehmen. Es handelt sich also um ein *dynamisches* Modell, das eine flexible Anpassung an die empirisch-operational (faktorenanalytisch) gewonnenen Befunde erlaubt.

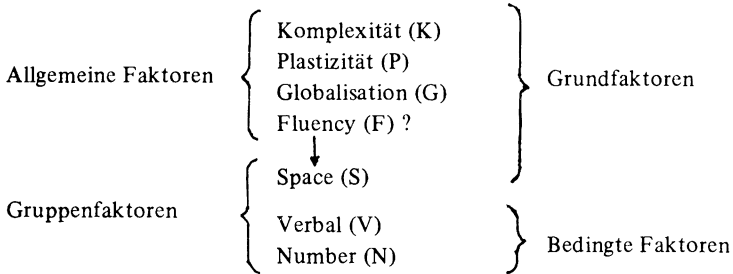
Von allen faktorenanalytischen Modellen der Intelligenz kommt dieser Entwurf der Englischen Schule – ausgenommen vielleicht den verwandten Meilischen Ansatz (siehe unten) – den phänomenologisch erarbeiteten Konzepten (vgl. Wenzl und Gottschaldt) am nächsten; zumindest sind gewisse Parallelen beider Ansätze auf der Interpretationsebene unverkennbar. Zugleich scheint damit ein interessantes Demonstrationsbeispiel dafür gegeben, daß ein fruchtbares Zusammenwirken von phänomenologischem (deskriptivem) und operationalem Ansatz keine Utopie zu sein braucht. Die gerade der Faktorenanalyse oft angekreidete Schwäche, an irgendeiner Stelle ihres Vollzugs ihre quantitativen Befunde interpretieren und in den jeweiligen Problemkontext einordnen zu müssen – die Notwendigkeit der Theoriebildung stellt sich auch für den puritanischen Operationalisten; genauso wenig kann der spekulative Denker der empirischen Veri- oder Falsifikation seiner Hypothesen entraten, will er sich nicht in wirklichkeitsfremdes (Wunsch-) Denken verlieren –, kann u.E. gerade in der Konvergenz beider Verfahrensansätze positiv aufgefangen werden. Siehe auch unsere früheren Ausführungen auf Seite 9 und 12f.

### *Das Faktorenmodell von R. Meili*

Meilis Intelligenzmodell, das von gestaltpsychologischen Theorien beeinflusst ist, enthält in der ersten Fassung (Meili 1944; vgl. auch 1961, S. 51 ff.) 4 *allgemeine Faktoren*, denen er „primäre Eigenschaften des Geistes“ zuschreibt. Es sind dies die Faktoren *Komplexität* (K), *Globalisation* (G), *Plastizität* (P) und *Flüssigkeit* (F). Diesen Gestaltungsprinzipien unterliegen nach Meili (auch) die intellektuellen Leistungsvollzüge bzw. im engeren Sinne die menschlichen Denkprozesse. Die Rolle der 4 Intelligenzfaktoren wird dadurch charakterisiert, daß sie völlig unabhängig voneinander variieren, wobei die „unabhängige Variation“ hier keine isoliert nebeneinander bestehenden „Fakultäten“ meint, sondern vielmehr andeuten soll, daß „mit jedem Ausprägungsgrad eines Faktors beliebige Ausprägungsgrade der übrigen verbunden“ sind, wodurch ein „inneres“ Abhängigkeitsverhältnis besonderer Art entsteht. M.a.W.: Bei Intelligenzleistungen sind immer alle 4 Faktoren im Spiele, aber eben zu jeweils wechselnden Anteilen; die unterschiedlichen Ausprägungsgrade resp. Stärkeverhältnisse scheinen von der Besonderheit der jeweils geforderten Intelligenzleistung (z.B. Anpassungsvermögen, Umstrukturierungsfähigkeit, Übersicht usw.) abhängig oder doch mitbedingt zu sein.

<sup>8</sup>Abk.: v:ed = verbal-numerical-educational; k:m = practical-mechanical-spatial-physical.

Die spätere Fassung der Meilischen Faktorentheorie enthält neben den 4 bzw. 3 bekannten allgemeinen Faktoren<sup>9</sup> noch die Kategorie *Gruppenfaktoren* (Fluency?, Space; Verbal und Number), die nicht generell, sondern nur an für sie jeweils spezifischen Denkprozessen beteiligt sind. Unter Berücksichtigung der Anlage-Umwelt-Antonomie teilt schließlich Meili seine Faktoren nach *Grundfaktoren* (vermutlich angeborene Eigenschaften) und „*bedingten*“ *Faktoren* (d.h. verstärkt soziokulturell bedingten, also erfahrungs- oder lernabhängigen Fähigkeiten) ein. Somit ergibt sich folgende Binnengliederung der Intelligenz sensu Meili (siehe auch Jäger 1967, S. 51):



Die 4 allgemeinen Intelligenzfunktionen lassen sich inhaltlich folgendermaßen bestimmen: *Komplexität* betrifft die „Ausdehnung des Feldes“, die „Reversibilität“ sensu Piaget bzw. die Fähigkeit zur Herstellung „sehr exakter Beziehungen zwischen zahlreichen verschiedenen Gegebenheiten“ nach Bergius; *Plastizität* meint die „Beweglichkeit“ oder Fähigkeit zur „Umstrukturierung“ und „Umzentrierung“; mit *Ganzheit* oder *Globalisation* ist die Fähigkeit zur Bildung von Ganzheiten und Ordnungen angesprochen; *Fluency* wird in Anlehnung an Thurstone als Leichtigkeit, von einer Idee zur andern hinüberzugleiten, interpretiert. Nach Bergius handelt es sich hierbei um die „Leichtigkeit der Produktion und Lockerheit der prästrukturalen Prozesse“, während Jäger auf einen wesentlichen Unterschied zum Faktor P aufmerksam macht: P betreffe die Fähigkeit zu Umformungen, bei F komme es dagegen auf „die Sicht anderer Seiten eines Gegebenen“ an, womit diesem Faktor F jedoch ein allgemeinerer Bedeutungsgehalt als bei Thurstone zugesprochen wird.

Der gleiche Autor verweist auf eine weitere wichtige Eigenschaft *sämtlicher* 4 Generalfaktoren Meilis. „Diese vier Faktoren (K, P, G und F) sind funktionaler Art, d.h. vom Material (sprachlich, numerisch, figural) der geforderten Leistungen unabhängig. Man kann sie als allgemeine Prozeßfaktoren bezeichnen, die die bei jedem Menschen (nur in verschiedenen Ausprägungsgraden) gegebenen Bedingungen („primären Eigenschaften“) repräsentieren, bestimmte Arten und Phasen von Denkprozessen zu bewältigen. Dabei wird angenommen, daß alle Denkprozesse eine einheitliche, allgemeinspsychologisch relevante Grundstruktur haben, die nur in ihren aktuellen Konkretisierungen in Abhängigkeit von der jeweiligen Aufgabe durch variierende Anteile dieser allgemeinen Prozeßfaktoren und ggf. durch die Mitwirkung von Gruppenfaktoren modifiziert wird“ (Jäger 1967, S. 52). Damit hat Meili – im Gegensatz zum mosaikartigen Strukturmodell von Thurstone – eine Brücke zur Denkpsychologie geschlagen.

Die restlichen (Gruppen-)Faktoren *Space*, *Verbal* und *Number* hat Meili offensichtlich von Thurstone (siehe unten) übernommen. Sie scheinen allerdings für Meili einen engeren Geltungsbereich zu haben.

<sup>9</sup>Zwischendurch rechnete Meili den 4. Generalfaktor (Flüssigkeit) den Gruppenfaktoren zu; in seiner letzten Veröffentlichung (1964, S. 139) findet sich Fluency wieder im Kreis der allgemeinen Faktoren, wobei ein Fragezeichen die schwankende Haltung des Autors bezüglich der endgültigen Zuordnung des Flüssigkeitsfaktors, dessen „Allgemeinheitscharakter nicht eindeutig klar“ sei, andeutet.

## *Die multiple Faktorentheorie von L.L. Thurstone (Modell der „Primärfähigkeiten“)*

Mit den folgenden Modellbeschreibungen knüpfen wir wieder an den ersten, von Spearman konzipierten Ansatz einer Zwei-Faktoren-Theorie der Intelligenz an. Nachdem eine Reihe späterer faktorenanalytischer Befunde mit der Hypothese Spearmans bezüglich des g-Faktors<sup>10</sup> nicht oder nur mit Hilfe von Zusatzannahmen in Einklang zu bringen war, begann Thurstone (1938, 1945; siehe weitere Literaturangaben bei Jäger 1967), die sog. Restvarianz, d.h. die nach der Extraktion von g verbliebenen *gemeinsamen* Varianzanteile, näher zu untersuchen (und nicht wie bisher als unbedeutende – weil nicht interpretierbare – Restvarianz zu vernachlässigen resp. auf unbekannte s-Faktoren zu verweisen). Dabei fiel ihm auf, daß man diese (Rest-)Varianzen durchaus psychologisch beschreibbaren Leistungseinheiten zuordnen konnte, was zugleich ihre sinnvolle Interpretation erlaubte. Mit Hilfe weiterentwickelter Techniken der von Spearman begründeten Faktorenanalyse, der sog. „multiplen“ Faktorenanalyse, entdeckte schließlich Thurstone eine Reihe von Grundfaktoren, die er als „Primärfaktoren“ der Intelligenz bezeichnete. Diese stehen in seinem Intelligenzmodell als völlig gleichberechtigte (also nicht hierarchisch strukturierte) Gruppenfaktoren nebeneinander und repräsentieren jeweils funktionale Einheiten von in sich ähnlichen Leistungsformen (homogenen Leistungsgruppen). Damit war nicht nur die Möglichkeit einer befriedigenderen Modell-(Theorie-)Zuordnung der faktorenanalytisch ermittelten Funktionsgrößen eröffnet, dieser Ansatz gewann zugleich plausible Bedeutung im Hinblick auf die konventionellen Alltagserfahrungen, wonach sich die Menschen nicht nur durch ihr Intelligenzniveau, sondern auch durch die Art und Anzahl ihrer Fähigkeitsrichtungen (z.B. mathematische Begabung, Sprachbegabung, technische Intelligenz usw.) unterscheiden. Die multiple Faktorentheorie Thurstones gestattet somit, die empirisch antreffbare Leistungsvielfalt intelligenten Verhaltens auf mehrere, gleichrangig nebeneinanderstehend gedachte Gruppenfaktoren zurückzuführen. Mit diesem Modell einer begrenzten Anzahl von gleichrangigen Gruppenfaktoren (die Thurstoneschen Angaben schwanken zwischen 7 bis 9 Primärfähigkeiten) setzte sich Thurstone einerseits gegen die Spearmansche Hypothese eines einzigen Grundfaktors (g) und andererseits gegen inflationäre Tendenzen anderer Vorstellungen, z.B. die unüberschaubare Vielzahl von Intelligenzfaktoren im Konzept Thorndikes, mit Erfolg ab. So überrascht es nicht, daß dieses Modell der Primärfähigkeiten wie kein anderes Faktorenmodell der Intelligenz die moderne Testkonstruktion entscheidend beeinflusst hat; die meisten jüngeren Intelligenztests basieren auf dem Thurstoneschen Entwurf. Dies hat nicht nur theoretische, sondern auch praktische Gründe. In der Intelligenz- und Eignungs-(Begabungs-)Diagnostik ist man sehr oft gezwungen, Begabungsklassifikationen (z.B. Differenzierung zwischen Gymnasial-, Real- und Hauptschuleignung versus zwischen Hauptschuleignung und Sonderschulbedürftigkeit (i.S. der Lernbehindertenpädagogik), zwischen der Eignung bzw. Begabung für den A- oder B-Kurs in Mathematik, Deutsch, Englisch usw.) vorzunehmen, wozu natürlich die theoretische Annahme einer Reihe voneinander *unabhängiger* Intelligenzen, also deutlich unterscheidbarer und damit klassifizierbarer Fähigkeiten, ein praktikables Modell zu bieten scheint. Inwieweit dieser Ansatz die diesbezüglich in ihn gesetzten Hoffnungen wirklich erfüllen kann, bleibt vorerst – trotz seines unübersehbaren Einflusses auf die Testpsychologie unserer Tage – abzuwarten. Die Praktikabilität ist noch kein Beweis für die empirische Gültigkeit des Modells. Thurstone selbst betonte (1938, S. 3 f.), daß es sich bei seinem Vorschlag nur um den „Versuch einer ersten Näherung“ handle. Und der Verfasser kann aufgrund seiner umfangreichen Erfahrungen in der Begabungsdiagnostik bzw. Schuleignungsprognostik diese vorsichtige Haltung nur bekräftigen. Freilich – das sei unmißverständlich ausgedrückt – kennen wir zur Zeit kein leistungsfähiges

---

<sup>10</sup>Spearman hatte ja ursprünglich angenommen, daß alle Intelligenzleistungen auf einen allgemeinen Faktor (general factor) zurückführbar und daneben nur spezifische Faktoren für jeweils eng umschriebene Spezialfähigkeiten verantwortlich seien.

ungsfähigeres Modell der Intelligenz als das von Thurstone<sup>11</sup>, und die Argumente pro und contra halten sich – in bezug auf die empirische Validierung des Faktorenansatzes – mindestens die Waage. Insofern wird man auch in der nächsten Zukunft Thurstones Faktorentheorie und ihre Aussagen bei der Begründung dessen, was intelligentes Verhalten ist und wie es funktioniert, zu berücksichtigen haben.

In der Literatur werden gewöhnlich 7 Primärfähigkeiten oder Gruppenfaktoren sensu Thurstone aufgeführt, deren Kenntnis im Zusammenhang unserer Problemerkörterung vorzugsweise im Hinblick auf die Interpretation von Intelligenz-Testbefunden interessant sein dürfte.

1. *Verbal Comprehension (V)*. Damit ist die Fähigkeit sprachlicher Bedeutungs- und Beziehungserfassung sowie der Umgang mit sprachlichen Begriffen gemeint. Dieser Faktor (oder richtiger: diese Faktorengruppe) repräsentiert nach allen Untersuchungsergebnissen eine der wichtigsten Funktionseinheiten im *menschlichen* Intelligenzbereich. Zwar spricht Thurstone die Vermutung aus, daß es weitere verbale Grundfaktoren(gruppen) gäbe, doch konnte bislang nur der folgende verbale Flüssigkeitsfaktor eindeutig ermittelt werden.

2. *Word Fluency (W)*. Hiermit ist die Leichtigkeit zu – relativ inhaltsunabhängigen – Wortverknüpfungen, also die mehr assoziative Wortproduktion angesprochen. Die Tatsache, daß es Personen gibt, die diskrepante Leistungen in V (wo die Wortbedeutungen eine Rolle spielen) und W (Mengenleistung der Wortproduktion) zeigen, läßt u.a. auf die Unabhängigkeit resp. Selbständigkeit beider Faktoren schließen. Einschlägige Erfahrungsbeispiele wären etwa der in der verbalen Assoziation (Wortproduktion) schwerfällige, gleichwohl das Werkzeug „Sprache beherrschende Denker auf der einen Seite und der „Schwätzer“ auf der anderen Seite.

3. *Memory (M)*. Dieser Faktor ist in erster Linie für mechanische Gedächtnisleistungen bzw. das Kurzzeitgedächtnis (Merkfähigkeit) kennzeichnend. Auch diese Faktorengruppe scheint sehr komplexer Natur zu sein, deren Korrelate für das beiläufige Erinnern genauso wie für das gezielte Memorieren, für Behaltensleistungen von sinnlosem wie sinnvollem Material verantwortlich sein dürften.

4. *Reasoning (R)*. Dieser vielleicht komplexeste Gruppenfaktor im Thurstoneschen Intelligenzmodell repräsentiert die Fähigkeiten zum logischen Schließen, zum Regel-Erkennen (Induktion) und zur Deduktion, womit die praktische Anwendung von Regeln oder Prinzipien angesprochen ist. Der Reasoning-Faktor (Denkfähigkeitsfaktor i.e.S.) scheint nicht an ein bestimmtes Aufgabenmaterial (sprachlich, numerisch, figural) gebunden zu sein.

5. *Number (N)*. Der numerische Faktor bezieht sich auf die Fähigkeit, relativ einfache Rechenoperationen, z.B. Additionen, Multiplikationen usw., auszuführen. Es handelt sich also hierbei weniger um Rechenfähigkeit als um Rechenfertigkeit. Mathematikbegabung im eigentlichen Sinne oder formallogisches Denken im Zahlenbereich werden hiermit kaum erfaßt. Thurstone selbst gibt zu bedenken, ob mit diesem Faktor nicht allgemeinere, d.h. nicht an Zahlenmaterial gebundene, Eigenschaften (traits) intendiert würden. Nach jüngsten Ergebnissen von Tent (1969, S. 96) wird mit diesem Faktor (N) wahrscheinlich eher die Leistungsmotiviertheit oder „anhaltende Konzentration bei geistiger Temporarbeit“ sensu Bartenwerfer erfaßt.

---

<sup>11</sup> Möglicherweise eröffnet das in der zweiten Hälfte der 50er Jahre vorgestellte Strukturmodell von Guilford neue, befriedigendere Verfahrensansätze. Eine empirisch nachweisliche Überlegenheit des Guilfordschen Klassifikationskonzeptes der Intelligenz war in bezug auf die Lösung begabungspsychologischer Probleme, wie sie etwa innerhalb des bestehenden schulischen Bildungssystems aktuell werden, bislang nicht zu belegen. Wir verweisen in diesem Zusammenhang nur auf die Erfolgskontrollen der sog. Kreativitätstests, die im Vergleich zu den traditionellen Intelligenztests überwiegend negativ ausfielen, was aber nicht notwendig gegen die neuen Tests spricht, sondern nur gegen deren Brauchbarkeit im derzeitigen Schul-Bildungs-System. Darauf sowie auf das Guilford-Modell werden wir weiter unten zurückkommen.

6. *Space (S)*. Der Räumlichkeitsfaktor repräsentiert das räumliche Vorstellungsvermögen und spielt – genau wie die Faktoren Closure 1 und 2 (siehe unten) – bei der technischen resp. technisch-konstruktiven Befähigung eine große Rolle.

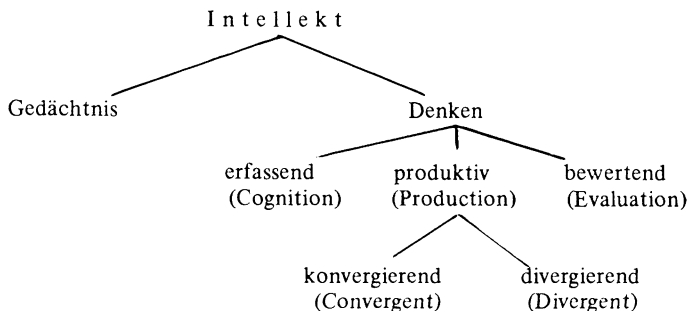
7. *Perceptual Speed (P)*. Diesen Faktoren wird die Fähigkeit zugeordnet, Details, die in irrelevantes Material eingebettet sind, rasch zu erkennen. Das Wahrnehmungstempo dürfte sowohl von der allgemeinen Auffassungsgeschwindigkeit oder Auffassungsleichtigkeit (Cognition) als auch vom individuellen Sehvermögen abhängen. Der Faktor P steht in gewissem Gegensatz zu Denkfaktoren i.e.S., die z.B. bei längeren Problemlösungsaufgaben oder bei der Verarbeitung von komplexem Material aktualisiert werden.

In späteren Veröffentlichungen setzte Thurstone für den Faktor P die Faktoren (8.) *Speed of Closure (Closure 1)*, der in enger Beziehung zu Meilis Faktor G (Globalisation) zu sehen ist, und (9.) *Flexibility of Closure (Closure 2)*, der Ähnlichkeit zu Meilis Faktor P (Plastizität) aufweist. Die beiden Faktoren Closure 1 und 2 möchte Thurstone unabhängig zu S (Space) interpretiert wissen.

Damit wäre das Bündel von Primärfähigkeiten im Thurstoneschen Strukturkonzept der Intelligenz formalinhaltlich umrissen. In unserem testpsychologischen Befundbericht bzw. im empirischen Teil werden wir auf das Thurstonesche Modell zurückkommen.

### *Das Intelligenz-Strukturmodell ("Structure of Intellect") J.P. Guilfords*

Die faktorenanalytische Forschung hat bisher rd. 100 Einzelfaktoren der Intelligenz ermitteln können. Guilford (1956, 1959a bzw. 1965), der in seinem z.Zt. wohl umfassendsten Klassifikationsmodell diese fast unüberschaubare Vielfalt von Einzelfaktoren systematisch zu ordnen versucht, differenziert die intellektuellen Fähigkeiten (Denkoperationen) zunächst nach der (kleineren) Gruppe der *Gedächtnisfaktoren* und der (größeren) Gruppe der *Denkfaktoren*. Letztere werden unterteilt in die drei Gruppen der *kognitiven* (erfassenden), *produktiven* und *wertenden* Denkfaktoren; das produktive Denken gliedert sich schließlich in die *konvergierenden* und die *divergierenden* Denkprozesse. Folgende Schemazeichnung möge das soeben Ausgedrückte verdeutlichen:

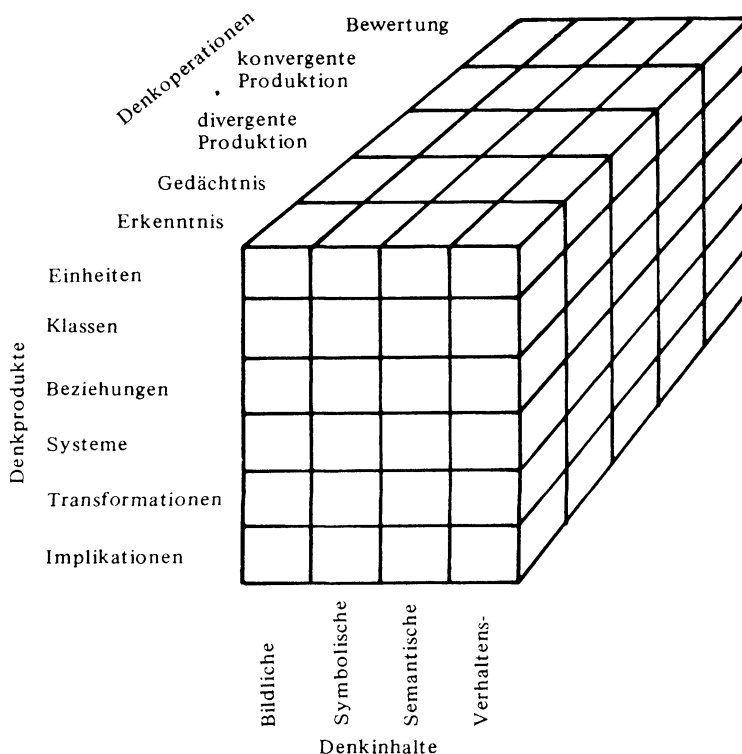


„Die *kognitiven* Fähigkeiten haben mit dem Auffinden von Information, mit Wiederentdecken und Wiedererkennen zu tun. Die *produktiven* Fähigkeiten betreffen die Anwendung bekannter Information, bisweilen auch, um dabei neue Information zu gewinnen. *Bewertende* Fähigkeiten treten auf den Plan, wenn es gilt, zu entscheiden, ob die erkannten Dinge oder entdeckten Sachverhalte richtig zusammenpassen, den Anforderungen genügen oder sonstwie adäquat sind... Das *konvergente* Denken zielt in eine einzige Richtung, man sucht eine herkömmliche (richtige ) Antwort oder schöpft eine einzige neue Lösung eines Problems. Beim *divergenten* Denken schreitet man dagegen nach verschiedenen Richtungen fort, beispielsweise bei der forschenden Tätigkeit, wo das Denken oft wechselnde Wege beschreiben muß“ (Guilford 1965, S. 353; kursiv v. Verf.).

Besonders interessant für die allerjüngste bzw. kommende Intelligenzforschung scheint die Guilfordsche Unterscheidung von konvergierendem und divergierendem Denkverhalten zu sein. Während mit *konvergierendem* Denken das „Herausarbeiten logischer Beziehungen“ und „das Ordnen von Begriffen“ gemeint ist, sind im *divergierenden* Denken „alle Maße der assoziativen und sprachlichen Flüssigkeit und Flexibilität sowie Umstrukturierungen und relativ freie Ergänzungsleistungen (Lückentests)“ betroffen (Graumann 1969, S. 41 f.). Unter traditionellen Klassifikationsvorstellungen könnte man hier von „gebundenem“ (kontrolliertem, systematischem) Denken – Convergent Thinking – und von „freiem“ Denken (frei assoziierendem Denken oder Phantasieren = „Vorstellungsablauf“) sprechen. Daß es offenbar auch Verlaufsmodi gibt, die „dazwischen“ liegen, also frei *und* gesteuert sein können, zeigt die Fähigkeit zum plötzlichen „Umstrukturieren“ (i.S. der Gestaltpsychologie), beispielsweise im Problemlösungsverhalten, worauf Graumann (loc. cit.) hinweist.

Die oben genannten 5 Intelligenzfunktionen werden weiterhin nach dem *Inhalt* oder Materialaspekt sowie der Art der Denk*produkte* aufgegliedert. Zu den *Denkinhalten* gehören *figurale* (jedes konkrete, wahrnehmbare Material, z.B. visuelle, akustische, haptisch-taktile, kinästhetische Vorstellungen), *symbolische* (Zahlen, Buchstaben bzw. Silben, „bedeutungsfreie“ Wörter und Sätze), *semantische* (sprachliches Material *mit* Bedeutungs-

Theoretisches Modell für die Gesamtstruktur des Verstandes (Theoretical model for the complete „Structure of Intellect“, 1962). Mit freundl. Genehmigung des Beltz Verlages entnommen aus: J. P. Guilford, Persönlichkeit. Weinheim 1965, 2/3. Aufl., S. 388.



gehalt, z.B. verbale Aufgaben, zu deren Lösung Bedeutungserfassungen notwendig sind), *behaviorale* oder soziale Inhalte („soziale Intelligenz“ bzw. „soziale Einfühlung“), deren faktorielle Bestimmung operational bis jetzt allerdings noch nicht gelungen ist.

Unter der Kategorie der *Denkprodukte* differenziert Guilford zwischen *Units, Classes, Relations, Systems, Transformations* und *Implications* (siehe Modellskizze).

Mit Hilfe der aufgezeigten 3 Dimensionen konstruierte Guilford sein bekanntes Würfelmodell, das insgesamt 120 (theoretisch postulierte) Intelligenzfaktoren enthält. Dem Modell eignet nicht nur eine bestechende Ordnungsfunktion, es besitzt auch heuristischen Wert, Ob allerdings eines Tages *alle* 22 Leerzellen des Würfels (98 Faktoren wurden ja bereits ermittelt) gefüllt sein werden, bleibt mehr als zweifelhaft. Guilford selbst betont die „logische“ Struktur seines Modells, was (sensu Jäger 1967, S. 97) „vermutlich zum Ausdruck bringen will, daß sein Entwurf zwar in sich folgerichtig, widerspruchsfrei und subjektiv evident, aber seine Übereinstimmung mit der Empirie damit noch nicht notwendig gegeben bzw. erwiesen sei.“ Immerhin konnte seit der ersten Modell-Veröffentlichung 1956 inzwischen eine Anzahl weiterer Intelligenzfaktoren entdeckt werden, was die heuristische Funktion des theoretischen Ansatzes unterstreicht (vgl. Guilford & Hoepfner 1971).

### Exkurs: Ergebnisse der modernen Kreativitätsforschung

Die Bedeutung des Guilfordschen Klassifikationskonzeptes für die Intelligenzforschung wurde schon hervorgehoben. Seine theoretische und praktische Relevanz erweist sich vielleicht am augenfälligsten in den Fortschritten jener jungen Richtung, die unter der Bezeichnung „Creativity“-Forschung in die Literatur Eingang gefunden hat. Der theoretische – und faktorenanalytische – Ansatz Guilfords trug wesentlich zur Systematisierung der in der Kreativitätsforschung bislang (spekulativ und empirisch) zutage geförderten Befunde bei.

Kreatives Verhalten („schöpferisches Denken“) scheint demnach in engster Beziehung zu den DP-Faktoren (Divergent Production) der „Operation“-Kategorie zu stehen. Obwohl bei Erwachsenen als auch bei älteren Schülern konnte Guilford (1964) 8 solcher DP-Faktoren abgrenzen: <sup>12</sup> DSU (Wortflüssigkeit), DMC (Semantische spontane Flexibilität, besonders bei männlichen Probanden), DMU (Gedankenflüssigkeit, besonders bei weiblichen Probanden), DMR (Assoziative Flüssigkeit, besonders bei weiblichen Probanden), DSS (Divergentes Produzieren symbolischer Systeme, stärker bei weiblichen Probanden), DFT (Figural adaptive Flexibilität), DMT (Originalität, stärker bei männlichen Probanden) und DMI (Elaboration). <sup>13</sup>

Korrelationen zwischen DP- und IQ-Tests (traditionellen Intelligenztests) ergaben Zusammenhangswerte zwischen  $-0.20$  und  $+0.52$ . Die Regressionslinie der IQ auf die DP-Werte verläuft nichtlinear (bei negativer Akzeleration). Am ehesten ist noch eine lineare Korrelation im unteren und mittleren IQ-Bereich festzustellen, während bei IQs über 120 keine Beziehung zwischen IQ- und DP-Werten ermittelt werden konnte. Dies bedeutet, daß DP-Tests möglicherweise adäquatere Diagnostika in den oberen Intelligenzbereichen (z.B. für die Erfassung der Real- und Oberschulbegabung bzw. der Studi ereignung) darstellen als die bislang verwendeten Intelligenztests. Eine solche generalisierende Interpretation der operationalen Befunde bedarf freilich noch eingehender Validitätsuntersuchungen, wobei Studien- resp. Schul- und Berufserfolg die „aka-

---

<sup>12</sup>Der folgende Abschnitt ist unserem Sammelbandbeitrag „Hochschüler-Guidance“ (vgl. Heller, Demel und Schorre 1969, bes. S. 140-143) entnommen.

<sup>13</sup>Siehe Modellskizze auf Seite 20. – *Abk.*: D = Divergent Production; F, S, M = Contents (Figural, Symbolic, Semantic); U, C, R, S, T, I: P = Products (Units, Classes, Relations, Systems, Transformations, Implications).



demischen“ Qualifikationskriterien repräsentieren müssen. Streng genommen beinhalten die Untersuchungsergebnisse nur, daß die traditionellen IQ-Tests und die neuen DP-Tests (relativ) Verschiedenes messen.

Der Katalog von *Wesenseigenschaften schöpferischer Persönlichkeiten* läßt an Heterogenität nichts zu wünschen übrig, was sicher nicht (nur) mit dem Hinweis auf die vermutete Heterogenität des Phänomens „Creativity“ erklärt werden kann. Die wissenschaftstheoretischen Unterschiede der einzelnen Forscherstandpunkte einschließlich der Diskrepanz ihrer Untersuchungsmethoden dürften kaum weniger an dem Konglomerat von Meinungen und Fakten beteiligt sein. So führt etwa Anderson (1959) am Schluß seines Buches über Creativity folgende Merkmale auf: den Wunsch zu wachsen, die Fähigkeit erstaunt zu sein, Aufmerksamkeit, Spontaneität, spontane und adaptative Flexibilität, Originalität, divergentes Denken, Offenheit für neue Erfahrungen, Durchlässigkeit von Grenzen, Hartnäckigkeit, Differenzierung, Integration, im Einverständnis mit der Welt leben, Harmonie, Bescheidenheit, Glaube, aber auch Enthusiasmus, Kühnheit, Skepsis, Mut zu einem vorübergehenden Chaos u.a.m. Die Meinungen der verschiedenen Autoren zusammenfassend konstatiert Anderson: Creativity ist das Leben selbst (Sinnott), Creativity ist eine Art des Lebens (Dow), Creativity ist optimales Wachsen in sozialer Interaktion (Anderson), Creativity ist das Maximum der Selbstverwirklichung (Maslow). Ähnlich hoben Hilgard und Gough die Relevanz von *Persönlichkeitsvariablen* hervor, während Stoddard das gefühlsmäßige Engagement der kreativen Persönlichkeit betonte. May hingegen meint, man solle nicht von schöpferischen Persönlichkeiten, sondern besser von den schöpferischen Akten sprechen. Anderson spricht schließlich noch von sozial-integrativem Verhalten, womit er den doppelpoligen Prozeß zwischen dem Menschen und seiner Umgebung erfassen möchte. Gemeint ist hier die Begegnung des äußerst bewußten Menschen mit seiner Welt (May). Auch Rogers betonte, daß Creativity nicht auf bestimmte Inhalte und Gegenstände begrenzt sei. Vielmehr sei für den schöpferischen Prozeß das Sichtbar-Werden eines neuen Produktes in der Aktion kennzeichnend. Im Bemühen des Menschen, sich selbst zu verwirklichen und seine Fähigkeiten zu aktivieren, läge die Motivation zu kreativem Verhalten. Diese sei zwar jedem Individuum eigen, könne aber durch Abwehrmechanismen u.ä. verdeckt sein. Als innere Voraussetzungen konstruktiver Kreativität werden genannt: Erfahrungen gegenüber dem Geöffnet-Sein, also das Gegenteil von Abwehr bzw. das Fehlen von Rigidität; innere Wertschätzung von Objekten; Unabhängigkeit vom Urteil der Außenwelt; die Fähigkeit, mit Elementen und Konzepten zu spielen. Fromm schließlich zählt zu den Voraussetzungen die Fähigkeit, erstaunt und verwirrt sein zu können, die Fähigkeit zur Konzentration, die Ich-Erfahrung (d.h. das Subjekt erlebt sich bewußt als Mittelpunkt der Welt, als Initiator seiner Handlungen), die Fähigkeit, Konflikte und Spannungen zu ertragen, die als Quellen des Staunens gesehen werden.

Diesen – zugegeben nicht immer sehr systematischen – Kennzeichnungen der Attitudes oder Traits bezüglich kreativen Verhaltens stellt Guilford folgende Fähigkeiten (Aptitudes) zum schöpferischen Denken gegenüber, wobei seine Informationen vorwiegend aus einem an der University of Southern/Calif. durchgeführten Projekt „Aptitudes of highlevel personnel“ stammen. Guilford (1959, S. 142 ff. – siehe auch S. 21 in diesem Buch) benennt 6 Aptitude-Faktoren (sowie einige Non-aptitude Traits), die in Beziehung zur Kreativität stehen.

1. *Fähigkeitsfaktor, Probleme zu erkennen bzw. eine allgemeine Sensitivität Problemen gegenüber.* Dieser Faktor wird am besten durch Tests erfaßt, in denen Fehler oder Mängel bei gebräuchlichen Werkzeugen oder sozialen Einrichtungen festzustellen sind oder einschlägige Probleme erkannt werden sollen.
2. *Fluency Faktor.* Hiermit ist der Ideenreichtum, vorab unter quantitativem Aspekt, gemeint: Wortflüssigkeit, Assoziationsflüssigkeit, Gewandtheit des Ausdrucks und Gedankenflüssigkeit, Wortschatztests, Benennen von Begriffsinhalten u.ä. Aufgaben sind hier relevant.
3. *Faktor der Flexibilität des Denkens.* Guilford unterscheidet in diesem Zusammenhang zwischen spontaner Flexibilität (Fähigkeit, eine Vielzahl von unterschiedlichen Ideen

zu produzieren) und adaptativer Flexibilität (z.B. Dispositionen für ungewöhnliche Problemlösungen, wie sie etwa die bekannten Streichholzspiele oder die sog. Denksportaufgaben verlangen).

4. *Faktor der Neubestimmung (Redefinition)*. Hier geht es darum, die Interpretation bekannter Objekte aufzugeben, damit diese auf neue Weise gebraucht werden können. Eine entsprechende Testfrage würde z.B. lauten: Welche der folgenden Dinge oder ein Teil davon könnten am besten zur Herstellung einer Nadel dienen? (Bleistift, Schuh, Fisch, Nelke). Auch die Fähigkeit zur Improvisation kann hiermit erfaßt werden.
5. *Faktor der Ausarbeitung (Elaboration)*. Die Relevanz dieses Faktors zeigt sich etwa, wenn ein Plan nur im Umriß vorgegeben wird und der Proband nun alle Detailschritte aufzählen muß, die zum Funktionieren des Planes führen. Hierbei sind sowohl figurale als auch bedeutungsgeladene (inhaltliche) Fähigkeitsmomente, die in positiver Korrelation zueinander stehen, notwendig.
6. *Faktor der Originalität*. Hierunter fallen nichtübliche Antworten (größere Zahl „origineller“ Lösungen), außergewöhnliche Assoziationen oder Beziehungen und eine bestimmte Anzahl von Antworten, die als „clever“ eingestuft werden (z.B. Erfindung von klugen Titeln für Kurzgeschichten).

Überraschenderweise fand Guilford keine Analyse- und Synthesefaktoren im Denken sowie – in gewissem Gegensatz zu Hilgard – keinen Problemsolving-Faktor, d.h. hinsichtlich dieser Kriterien unterscheiden sich die kreativen Persönlichkeiten nicht von den anderen. Zu den *Non-aptitude Traits* zählt Guilford die *spontane und adaptative Flexibilität*, die Freiheit von Perseverationen bedeutet, d.h. Wegkommen von früher gelernten Lösungswegen. Beide Formen der Flexibilität wären hier im Gegensatz zum Rigiditäts-Faktor zu sehen.

Resümierend halten wir somit fest, daß die Ungewöhnlichkeit (Originalität) der Leistung, Ideenreichtum, eine besondere Fähigkeit zur Lösung unklarer Probleme, kognitive Beweglichkeit und Umstrukturierungsvermögen resp. soziale Anpassungsfähigkeit<sup>14</sup>, neben allgemeintellektuellen Fähigkeiten eine ausgeprägte Fragelust als Geisteshaltung sowie Zielbestimmtheit, z.T. auch ästhetisches Feingefühl (sensu Gough) und – quasi als „äußere“ Bedingungsvariablen – bestimmte soziale (familiäre und schulische) Hintergrundstrukturen die *Hauptcharakteristika schöpferischer Leistungsdisposition* darstellen. Einschlägige empirische Untersuchungen zur Kreativität, etwa die von Holland (1964) an einer ausgewählten Gruppe Hochbegabter (Stipendiaten), orientierten sich sub specie akademischer Leistungseffizienz bislang vorab an *Kriterien öffentlicher Anerkennung*, z.B. Preisen, Auszeichnungen, Publikationen. Hierbei fand man, daß der Highschool- bzw. Collegestudent, sofern er kreative Leistungen in den Natur- und Geisteswissenschaften aufweist, weitgehend dem Stereotyp des (*Natur-/Wissenschaftlers* und *Künstlers*) entspricht. Siehe noch das Sammelreferat von F. Barron (1965) sowie die deutschsprachige Einführung von G. Ulmann (1968).

Der Exkurs in die moderne Kreativitätsforschung erscheint vielleicht manchem Leser – in Anbetracht der zentralen Thematik dieses Buches – allzu umfangreich geraten. Im Hinblick auf die Bedeutung dieser jungen Forschungsrichtung, nicht zuletzt für die schulpädagogischen Anliegen des „Begabens“, sowie die lawinenartig anwachsende Fülle von Einzelarbeiten und Konzepten zum Thema „Creativity“, trägt unser Beitrag gleichwohl sehr aphoristische Züge. Für die eingehendere Beschäftigung mit den angeschnittenen Problemen, vorab unter schulpädagogischen Gesichtspunkten, empfehlen wir noch Massiales & Zevin (1969), Mühle & Schell (1970), Skowronek (1968); Einzelfragen werden im Kontext testtheoretischer bzw. psychologisch-diagnostischer Erörterungen erneut aufgegriffen (vgl. u.a. S. 75 u. 218). Schließlich sei das 1945 in erster Auflage erschienene, nach wie vor lesenswerte Buch von Wertheimer über „Produktives Denken“ empfohlen.

<sup>14</sup> Allerdings trifft Hilgard die bemerkenswerte Feststellung, daß gute soziale Anpassung nicht die beste Voraussetzung für alle Formen menschlicher Produktivität darstelle. Siehe hierzu Hilgard 1959, S. 162 ff.

Die jüngste und wohl umfassendste (deutschsprachige) Arbeit zur faktorenanalytischen Bestimmung der Intelligenz stammt von Jäger (1967). Nachdem wir bereits wiederholt auf diese Arbeit Bezug genommen haben, wollen wir zum Schluß unseres kleinen Sammelreferats über die Faktorentheorien den – vorerst als „Näherungslösung“ zu bewertenden – Jägerschen Modellentwurf erörtern.

Jägers Test-(Untersuchungs-)Material bestand aus insgesamt 289 (!) Einzelvariablen; davon konnten mit Rücksicht auf die Speicherkapazität des DRZ-Computers<sup>15</sup> IBM 7090 maximal 234 Variablen faktorenanalytisch (simultan) verarbeitet werden. Eine Korrelationsmatrix in dieser Größenordnung (234 x 234) dürfte selbst in den USA bis dato noch nicht faktorenanalysiert worden sein. Trotz der Extensität des Aufgabenmaterials (Tests), das sich vor allem von der Absicht her, „eine möglichst repräsentative Stichprobe intellektueller Leistungsformen zu erfassen“, <sup>16</sup> erklären läßt, konnte eines der Hauptziele des Autors, nämlich die Faktorenmodelle Meilis, Thurstones und Guilfords empirisch zu verifizieren, nur teilweise und in keinem Falle voll befriedigend erreicht werden. Sofern man den Grad der Affinität zu Jägers faktorenanalytisch ermittelten Hauptdimensionen der Intelligenz als Maßstab anlegt – 5 von Jägers 6 Hauptfaktoren zeigen eine mehr oder weniger deutliche Übereinstimmung (auf Interpretationsebene) zu den oben beschriebenen Primärfähigkeiten von Thurstone (1 = S, 2 = W? 4 = R, 5 = N, 6 = V?)<sup>17</sup> –, eignet dem Thurstoneschen Intelligenzkonzept von allen dargestellten Faktorenmodellen die größte Empirie-Wahrscheinlichkeit. Hingegen ist nach Meinung Jägers (a.a.O., S. 174 f.) Guilfords Würfelmodell „allzu 'logisch'-systematisch und zu unorganisch konzipiert“, was sich u.a. in der Schwierigkeit, dieses Klassifikationskonzept mit den Daten seines (Jägers) Testmaterials zu konfrontieren, manifestierte. Oder sollte methodische Unzulänglichkeit ein solches Vorhaben (vorerst noch) zum Scheitern bringen, d.h. das sehr komplexe Modell operationalem Zugriff sich entziehen? Immerhin werden die Ordnungsfunktion und der heuristische Wert des Guilfordschen Klassifikationsansatzes der Intelligenz von Jäger hoch eingeschätzt.

Neben der empirischen Kontrolle der Geltungsbereiche konkurrierender Faktorenmodelle galt Jägers zweite Zielfrage der *Erhellung der faktoriellen Intelligenzstruktur*, d.h. der Erfassung von Strukturkomponenten oder Faktoren, welche am wahrscheinlichsten umfassendere Intelligenz-Leistungsbereiche repräsentieren. In den zahlreichen Faktorenanalysen konnten 6 *Hauptdimensionen der Intelligenz* oder *Hauptfaktoren(gruppen)* abgegrenzt werden, die folgendermaßen interpretiert wurden.

---

<sup>15</sup> DRZ = Deutsches Rechenzentrum in Darmstadt

<sup>16</sup> Insgesamt wurden 419 Unter- und Oberprimaner an zwei Unterrichtstagen mit über 250 Einzeltests untersucht; nach Abzug der wegen besonderer Gründe (z.B. Unvollständigkeit der Protokolle) unauswertbaren Testunterlagen blieben schließlich 301 Probanden übrig, deren Material für die statistische resp. faktorenanalytische Verarbeitung benutzt werden konnte.

Die Homogenisierung der Stichprobe und die hieraus resultierende Einengung der Generalisationsbasis (streng genommen gelten die Untersuchungsergebnisse nur für Primaner) wurde vom Autor bewußt in Kauf genommen zugunsten der Repräsentation des Intelligenzleistungsfeldes, d.h. einer (erhofften) breiteren Faktorenstruktur. Allerdings wird die Zuverlässigkeit einer Reihe von Einzeltests durch extrem kurze Bearbeitungszeiten u.E. erheblich eingeschränkt, so daß man sich fragt, ob sich der immense Materialaufwand (speziell in diesen Fällen) lohnte.

<sup>17</sup> Siehe S. 18f. in diesem Buch. Zur Aufschlüsselung der Jägerschen Faktoren vgl. umseitige Tabellenübersicht, woraus die 6 Dimensionen oder Hauptfaktoren resp. deren Interpretationsvorschläge zu ersehen sind.

1 *Anschauungsgebundenes Denken*. „Es ist der am breitesten fundierte Faktor.“ (S. 114). Relevante Testaufgaben finden sich z.B. im *Hamburg-Wechsler-Intelligenztest für Kinder* (HAWIK) bzw. *für Erwachsene* (HAWIE), vorab im *Mosaik*-Subtest (wo mit Hilfe mehrfarbiger Klötze bildhaft vorgegebene Muster gelegt werden müssen), ähnlich bei der *Nicht-verbalen Intelligenztestreihe von Snijder-Oomen* (SON), oder im *Charkow-Test*, der besonders im Rahmen der Berufsberatung bei der Berufseignungsermittlung – auch bei Gehörlosen – oft Verwendung findet (nach bestimmten Regeln aufgebaute Figuren-Reihen sind logisch fortzusetzen bzw. zu ergänzen); auch der *Intelligenz-Struktur-Test* (IST) von Amthauer und das *Leistungs-Prüf-System* (LPS) von W. Horn enthalten entsprechende Aufgaben (Figuren- und/oder Zahlenmatrizen, die zu ergänzen sind bzw. in denen versteckte Fehler – gegen die Gesetzmäßigkeit des Aufbaus – herausgefunden und markiert werden müssen)<sup>18</sup>. Jäger interpretiert den Faktor „Anschauungsgebundenes Denken“ (funktional) als „Fähigkeit und Neigung zu anschauungsgebundenen Denkvollzügen ggf. auch unter Einschluß einer Transponierung andersartigen Materials in ein anschauliches Medium“. Ferner vermutet er, „daß infolge der Eigenart unserer Variablenstichprobe einige weitere Faktoren relativ hoher Allgemeinheit gerade ungeschieden in ihn eingegangen sind, und zwar visuelle Auffassungsgabe, allgemeine Auffassungsgabe und der bekannte Performance-Faktor, welcher die Praktische Intelligenz repräsentiert“ (S. 175). Zwar beinhaltet der Hauptfaktor 1 in stärkerem Maße räumlich-visuelle Vorstellungsfähigkeiten, doch lehnt Jäger eine *direkte* Identität mit Thurstones Faktor S (Space) ab, da sein Faktor „weitaus komplexer als Thurstones Space“ und stärker mit Reasoning-Anteilen (z.B. Regel-Erkennen) gesättigt sei. Unter diesem Gesichtspunkt ist sicherlich auch der *Progressive Matrizen-Test* (PMT) von Raven hier einschlägig, den wir später noch ausführlich in seiner Bedeutung für die Hörgeschädigtendiagnostik würdigen werden.

2 *Einfallsreichtum und Produktivität*. Auch dieser Faktor ist breit fundiert und erweist sich als „hochgradig konsistent“. Dies ist um so interessanter, als der hier beschriebene Faktor 2 in der Intelligenztheorie faktorenanalytischer Provenienz (ausgenommen Guilfords Modell), besonders aber in der praktischen Intelligenzdiagnostik via Tests, mehr als stiefmütterlich behandelt wurde. Einschlägige Testaufgaben sind etwa das Erfinden bzw. Auswählen von cleveren Kurzgeschichten, Überschriften, Titel-Bildern, originellen Zeichnungen usw. oder das Fällen von Entscheidungsurteilen anhand von Eigenschafts- und Fähigkeitslisten (z.B. Herausfinden der für einen bestimmten Beruf, etwa Techn. Zeichner, Friseur, Automechaniker, Verkäufer(in) im Warenhaus etc., hinderlichen Eigenschaften und Fähigkeiten). Auch andere Verfahren, wie der sog. Konsequenz-Test bei dem möglichst vielseitige Folgerungen aus unerwarteten Situationen gezogen werden sollen (z.B.: Welche Folgen hätte es, wenn plötzlich aus allen Wasserleitungen der Stadt Salzwasser flösse?), oder der Insight-Test von Sargent, dessen 15 Aufgaben diverse Konflikte und Probleme menschlicher Anpassung – im Hinblick auf Familie, Freundschaften, anderes Geschlecht, Gesundheit, Meinungsbildung, Religion, Beruf usw. – symbolisieren (z.B.: Ein junger Mann erfährt vom früheren, schlechten Lebenswandel seiner Freundin. Wie verhielt er sich? Was hat er empfunden?), scheinen in enger faktorieller Beziehung zu Einfallsreichtum und Produktivität zu stehen.

Der Faktor 2 zeigt auffallende Ähnlichkeit zum *Fluency*-Faktor Meilis, der weiter als der (nur die „Wortflüssigkeit“ intendierte) Fluency-Faktor Thurstones interpretiert wird. Am überzeugendsten ist jedoch die Übereinstimmung mit Guilfords *Divergent Production*, wozu wir bereits im Rahmen unseres Creativity-Exkurses wichtige Beschreibungs- und Interpretationsmerkmale lieferten.

3 *Konzentrationskraft und Tempo-Motivation (besonders bei einfach strukturierten Aufgaben)*. „Faktor 3 kommt vornehmlich bei leichten Routineaufgaben ins Spiel, wo angespannte Konzentration und rasche Auffassung gefordert werden und das Resultat vornehm-

<sup>18</sup> Die meisten der hier genannten Tests werden im empirischen Teil des Buches eingehender beschrieben. Vgl. Seite 114ff.

# Faktorenstruktur der Intelligenz nach A.O. Jäger\*

(Nr.) Hauptfaktor/ Intelligenzdimension (sensu Jäger 1967)	Engere Zugehörigkeit der spezielleren Faktoren zu dem betreffenden Hauptfaktor	Vermutlich bereichsfremde oder weniger eng zugehörige spezielle Faktoren	Hinweis auf gleiche oder ähnliche Faktoren, die in anderen Binnenstrukturanalysen aufgewiesen wurden	Vermutete Beziehungen zu ähnlichen Faktoren	Vermutete Beziehungen zu Bereichen und Eigenschaften außerhalb der konventionellen Intelligenzfaktoren
1 Anschauungsgebundenes Denken	Visualization	Cognition (Auffassungs-leichtigkeit) (siehe unten)			Wahrnehmungs- und Persönlichkeits-eigenschaften
	Perceptual Foresight			4	
	Figural Adaptive Flexibility			2	
	Produktion bei figuralem Material	Produktion bei Zahlenmaterial	2		
2 Einfallsreichtum und Produktivität	Originalität		2	4	Persönlichkeitseigenschaften und Motivation
	Sprachl. Material:		6		
	a) Assoziationsflüssigkeit			6	
	b) Conceptual Foresight (Problem-sensitivität)			4, 6	
	Verbal Fluency		3	6	
	Zahlenmaterial:				
	a) Einfach strukturiertes		5		
	b) Komplexes		5		
	Figurales Material		1		
	Konzentrationskraft		3		
3 Konzentrationskraft und Tempo-Motivation	Zahlenmaterial	Produktion bei Zahlenmaterial	5		Motivation, Wahrnehmung und Persönlichkeitseigenschaften
	Clerical Material		2		
	Sprachl. Material		6		
	Merkfähigkeit		M		

4 Verarbeitungskapazität, formallogisches Denken und Urteilsfähigkeit	Judgment, Urteilsfähigkeit		2	Persönlichkeitseigenschaften
	Sprachl. Material	6		
	Komplexes Material	5	1, 6	
	Anschauungsgebundenes Denken	1		
	Konzentrationskraft	3		
5 Zahlengebundenes Denken	Verarbeitungskapazität	4		
	Einfallsreichtum	2		
	Produktivität	2		
	Konzentrationskraft	3		
	Figural Adaptive Flexibility	1		
6 Sprachgebundenes Denken	Symbolic Manipulation und Merkfähigkeit	M		
	Education of Semantic Relations	4		
	Originalität	2	4	
	Education of Semantic Correlates (Wortschatz? )			
	Verstümmelte Wörter		1, 2	
	Merkfähigkeit: sprachl.Material	M		
	Merkfähigkeit: Zahlenmaterial	M, 5		
<hr/>				
M Merkfähigkeit		3, 5, 6	1, 4	
Cognition (Auffassungs-leichtigkeit)	Anschauungsgebundenes Denken			
Performance, „Praktische Intelligenz“	(Hauptfaktor 1)			
Gedächtnis			1	Motorik
			1, 2, 4, 5, 6	

\*Unsere tabellarische Synopse stellt eine Vereinfachung der wesentlich anschaulicheren Modellzeichnung Jägers (a.a.O., S. 179/180) dar. Vor allem kommt die Verflechtung der Binnenstrukturen der Hauptfaktoren in der Originaldarstellung besser zum Ausdruck.

Die unter dem Strich aufgeführten Intelligenzdimensionen stellen von Jäger vermutete „weitere Hauptfaktoren“ höherer Allgemeinheitsgrade im konventionellen Leistungsbereich dar; dies betrifft besonders den (vorerst nur theoretisch) postulierten Hauptfaktor „Gedächtnis“.

lich von der Tempo-Motivation, vielleicht auch von Komponenten wie Stress-Resistenz, Störbarkeit u.ä. abhängig ist. Ob man ihn als Intelligenzfaktor oder als allgemeinen Leistungsfaktor ansehen will, ist eine Frage zweckmäßiger Bereichsabgrenzungen, die sich nur in weiteren Untersuchungen klären läßt" (S. 176). Einschlägig sind hier die gängigen Konzentrations-Meßverfahren, wie sie sich auch beim HAWIK bzw. HAWIE (Zahlen-Symbol-Subtest) oder im LPS (Arbeitsprobe) finden. 1961 hat E. Martin am Heidelberger Institut für Hör-, Sprach- und Sehgeschädigtenpädagogik Untersuchungen mit dem *Pauli-Test*, dem *Konzentrations-Verlaufs-Test* von Abels und dem *Figuren-Durchstreich-Test* von Bourdon an hörgeschädigten Kindern durchgeführt, über deren Ergebnisse wir ebenfalls im empirischen Teil berichten wollen.

In den Hauptfaktor 3 gehen also Variablen wie Konzentrationsfähigkeit, Lern- und Leistungsmotivation, aber auch Wahrnehmungseigenschaften sowie Auffassungs- und Materialkomponenten ein. Im Sinne der Mierkeschen Begriffsdifferenzierung haben wir es hier mit den sog. Hilfs- oder Stützfunktionen der Intelligenz zu tun.

**4 Verarbeitungskapazität, formallogisches Denken und Urteilsfähigkeit.** Sensu Mierke ist mit diesem Faktor die eigentliche Kernintelligenz getroffen. Aufgaben zum schlußfolgern des Denken, Wort- und Figurenanalogien, aber auch mechanisch-technisches Verständnis erfordernde Testitems oder eingekleidete Rechenaufgaben sind auf diesem Faktor gesättigt.

Der Faktor 4 zeigt gute Übereinstimmung zu Meilis Faktor *Komplexität* und zum *Reasoning*-Gruppenfaktor von Thurstone. Im Jägerschen Analysematerial fanden sich weitere Affinitäten zu Guilfords Faktorenklassen *Cognition* und *Evaluation*.

**5 Zahlengebundenes Denken.** Hinsichtlich der Funktion sind die Testaufgaben, die auf diesem Faktor laden, recht heterogen. Ihr gemeinsames Merkmal ist die Zahlengebundenheit des Materials. So sollen etwa möglichst viele Zahleneinheiten (z.B. Telefon-Nummern) erfunden bzw. zusammengestellt und gleichzeitig durch Beachten gewisser Konstruktionsprinzipien (Regeln) das Behalten der Zahlen erleichtert werden. Oder es sind in einer Kolonne alle jene Zahlen zu markieren, deren Abstand zur vorausgehenden Zahl durch eine bestimmte Konstante ausgezeichnet ist. In Jägers Aufgabenmaterial finden sich ferner Zahlenrätsel sowie eingekleidete Rechenaufgaben (Dreischluß-Textrechnungen).

Die Faktorenanalyse der aufgezählten Tests erbrachte Ähnlichkeiten des Faktors 5 zu Thurstones Faktor N, wenngleich *Number* dort viel enger gefaßt wird (mehr im Sinne bloßer Rechenfertigkeit in bezug auf die Grundrechnungsarten), zu Guilfords *kognitiven* (X-Konstante) und *bewertenden* (bei Schätzungsaufgaben) Faktoren. Aber auch zu den übrigen Funktions-Klassen Guilfords, besonders zum Faktor *divergente Produktion* (eingekleidete Rechenaufgaben) zeigten sich Übereinstimmungen, wohingegen sub specie Materialkategorie sämtliche kennzeichnenden Testleistungen in die Klasse *Symbolic* fallen.

**6 Sprachgebundenes Denken.** Der verbale Charakter des Aufgabenmaterials scheint das durchgehende Kennzeichen für diesen Faktor zu sein, dessen Identifizierung dadurch erschwert wurde, daß die Lösung verbaler Aufgaben immer mehrere Faktoren zugleich beanspruchte, besonders Faktor 1, 3 und 4. Mit Ausnahme der englischen Schule, die eine „gewisse Gemeinsamkeit aller sprachgebundenen Leistungen im Sinne eines Gruppenfaktors höherer Ordnung“ annimmt, scheinen die Ergebnisse der früheren Faktorenanalysen allesamt den Befund Jägers (Aufspaltung des verbalen Materials in verschiedene Faktoren) zu stützen. Evident wird in diesem Zusammenhang besonders Jägers vierte Erklärungshypothese, wonach „der Bereich der sprachgebundenen Leistungen im Vergleich mit den zahlengebundenen und den Leistungen bei figuralem Material (als) differenzierter“ anzusehen sei, was indirekt durch die „vielfältige, alle Lebensbereiche durchdringende Bedeutung der Sprache speziell bei höheren Bildungsstufen, ihr Primat als Kommunikationsmittel u.a.m.“ gestützt werden dürfte (S. 134).

Einschlägig sind demnach sprachlich in irgendeiner Form gebundene Testmaterialien, im Falle Jägers die bereits oben erwähnten Verbaltests, ja teilweise sogar Aufgaben wie „Zahlenmerken“, „Figurenmerken“ u.ä. „Den Signierungen nach entspricht der Faktor 6 keiner der faktoriellen Hypothesen Thurstones, Meilis oder Guilfords. Speziell mit Thurstones V (Verbal Comprehension; d. Verf.) ist er kaum zu identifizieren. Die verschiedensten Funktionen sind in den kennzeichnenden Leistungen vertreten, *kognitive* wie Verstümmelte Wörter (VW), *produktive* wie Gegensätze (GS) und Titel-Text (TT), *bewertende* wie Sprichwörter (SW) und Schlüsse (SL) und *Merkfähigkeit* wie Namenmerken (NM)“ (S. 133).<sup>19</sup>

Zu den aufgewiesenen 6 Hauptdimensionen der Intelligenz (siehe noch S. 26f.), deren Geltungsbereiche an Hand ausgedehnterer Populationsfelder und ggf. andersartiger Materialien zu überprüfen Jäger vorschlägt, gesellen sich weitere, vom Autor vermutete, jedoch empirisch noch nicht hinreichend gesicherte „Faktoren ähnlich hoher Allgemeinheitsgrade“ (besonders M = Merkfähigkeit, Performance, Kognition und – vorerst rein hypothetisch – Gedächtnis). Auf der anderen Seite müssen „Speziellere Funktionseinheiten“ angenommen bzw. die Möglichkeit zur weiteren Aufdifferenzierung der Binnenstrukturen als wahrscheinlich angesehen werden. Für künftige Forschungsansätze empfiehlt deshalb Jäger als Bezugsrahmen ein *hierarchisches* Modell analog der bei Burt und Vernon entwickelten Konzeption – allerdings *ohne* die Ausgangshypothese eines g-Faktors i.S. eines „vorwegnehmenden Postulates“. Freilich: „Sollte er sich bei empirischen Strukturanalysen, welche außer Intelligenzleistungen auch andere Verhaltensbereiche einbeziehen, ergeben, wird man ihn in ein dementsprechend umfassenderes Netz einordnen und durch Analyse seiner Binnenstruktur auch seine Beziehungen zu den Intelligenzfaktoren geringerer Allgemeinheitsgrade klären können“ (S. 182). Offensichtlich will damit Jäger die Berechtigung zur Annahme eines solchen Generalfaktors, der als Bedingungsfaktor aller Intelligenzleistungen in Frage käme, nicht grundsätzlich in Abrede stellen.

Vordringlicher ist zunächst jedoch die „*Integration von Intelligenz-, Motivations- und Persönlichkeitsforschung* mit dem Fernziel einer Verhaltenstheorie, welche sowohl die hier genannten nächstliegenden, als auch alle weiteren, bislang weitgehend getrennt erforschten Verhaltensbereiche einschließt. Die bisher herausgearbeiteten Intelligenzdispositionen sind quasi „conditioned factors“ (sensu Cattell). Bei ihrer Identifizierung unterstellt man, daß außer den variierten Aufgabenstellungen alle inneren und äußeren Leistungsbedingungen konstant gehalten wurden und durch eine optimale Aktivierung der Probanden die maximale Kapazität ihrer intellektuellen Fähigkeiten erkennbar wird. Bei der *theoretischen* Erfassung intelligenten Verhaltens geht man oft, wenn nicht meist, so vor, als ob die Intelligenzdispositionen die allein maßgebenden Bedingungen seien und motivationale, Persönlichkeits- und andere Bedingungen vernachlässigt werden könnten. Diese Annahme bedarf sicher der Revision“ (S. 184; kursiv v. Verf.). Daß diese Erkenntnis sich heute immer mehr durchsetzt, ist weniger ein Verdienst der Intelligenzforschung i.e.S. als der modernen Begabungsforschung<sup>20</sup>. Siehe auch Wewetzer (1972).

---

<sup>19</sup> Die Abkürzungen beziehen sich auf das von Jäger benutzte Aufgabenmaterial.

<sup>20</sup> Gewöhnlich findet man die Forschungsrichtungen, welche die „nichtintellektuellen“ (persönlich-motivationalen bzw. sozio-kulturellen) Determinanten intellektuellen Leistungsverhaltens akzentuieren, unter dem Stichwort „Begabungsforschung“ heute zusammengefaßt. Demgegenüber beschäftigt(e) sich die Intelligenzforschung in ihren empirisch-operationalen Ansätzen, also auch die Faktorenanalyse, vorwiegend mit der Intelligenz i.e.S. Diesen Gegenstandszentrierungen entspricht unsere frühere



Die These von der sozio-kulturellen (Mit-)Bedingtheit von Intelligenz und Begabung beinhaltet in ihrem pädagogischen und bildungspolitischen Aspekt geradezu ungeheuerliche Konsequenzen. Leider artet die Diskussion um das Für und Wider einer solchen Annahme-Berechtigung heute sehr oft in nutzlose Ideologie-Streitigkeiten aus. Um so notwendiger ist eine objektive, vorurteilsfreie Überprüfung dieser Frage. Eine knappe Bestandsaufnahme der bisher zu diesem Problemkomplex vorliegenden *empirischen* Untersuchungsbefunde soll uns dazu die notwendigen Informationen liefern, ohne die u.E. keine wirklichkeitsangemessene Entscheidung getroffen werden kann.

## 2. Die Frage nach der sozio-kulturellen Bedingtheit der Intelligenz im Lichte der neueren Begabungsforschung

Die Frage, ob bzw. in welchem Rahmen und in welchem Ausmaß die Dispositionen zu intelligentem Verhalten (gemessene Intelligenzleistung versus Lernleistung) eine gewisse *intraindividuelle Konstanz* aufweisen, interessiert vorab unter zwei Aspekten: einmal im Hinblick auf die Möglichkeit *zuverlässiger* Intelligenzmessungen und zum andern im Hinblick auf die Möglichkeit pädagogischer Einwirkung im Sinne des „*Begabens*“ oder – allgemeiner – hinsichtlich der Bildungschancen in bezug auf die individuelle Intelligenz- und Begabungsförderung. Diese Frage berührte uns schon einmal (indirekt) im Zusammenhang mit der Erörterung von Anlage/Erb- und Umweltanteilen beim menschlichen Intelligenzaufbau. Wohl am umfassendsten wurde diese Problematik neuerdings von Bloom (1966) dargestellt, der die wichtigsten Forschungsergebnisse, soweit sie für die *allgemeine* Intelligenzentwicklung Bedeutung erlangen, berücksichtigte (vgl. auch Lückert 1969b).

---

### Fortsetzung Fußnote 20

Differenzierung der Bedeutungsakzente von Intelligenz und Begabung (vgl. S. 10 ff., bes. S. 11, Fußnote).

Neben den unterschiedlichen Objektperspektiven lassen sich jedoch auch *methodische* Gründe für die disparate Einstellung beider Richtungen anführen. So war (und ist) es das Ideal aller experimentellen (und dazu gehören die psychometrischen, insbesondere intelligenzdiagnostischen) Designs, die zu untersuchende Variable (in unserem Falle die Intelligenz) durch Kontrolle, d.h. Konstanthaltung der übrigen Bedingungskomponenten (hier der Umwelt- und Persönlichkeitsvariablen) sicher in den Griff zu bekommen (vgl. obiges Zitat). Dagegen ist die Pädagogische Psychologie (resp. der Pädagoge und Erzieher) natürlich viel stärker an den Möglichkeiten des „*Begabens*“, also den motivationalen und sozialen (schulischen und familialen) Determinanten der Begabungsentwicklung versus der Intelligenzdiagnose im konventionellen Sinne qua Grundlage und Voraussetzungsbedingung der Begabungsförderung des Schülers interessiert. Insofern stehen der psychologische (operational definierte) Begriff der Intelligenz und der mehr pädagogisch gesättigte Begriff der Begabung in gewissem Widerstreit. Die Notwendigkeit zur Konvergenz beider Sichtweisen in Richtung eines umfassenderen Intelligenzkonzeptes (=Begabung), das den schul- und berufspraktischen Anforderungen besser als bisher genügen kann, wurde schon des öfteren betont. Die Gefahren für die operationale *Intelligenzforschung* liegen vorab in der Verengung ihres Blickwinkels bezüglich des Bedingungskomplexes „Intelligenz“, die Hauptgefahr der modernen *Begabungsforschung* besteht in der Aufweichung ihres Begabungs-(Intelligenz-)Begriffs resp. der vagen, sehr oft empirisch nicht zu stützenden Bestimmung der Intelligenz- und Begabungsinhalte (Intelligenzdimensionen oder -faktoren). Sofern sich jedoch beide Forschungsansätze jeweils der Aspekthaftigkeit ihres Vorhabens bewußt sind – die ermutigenden Anzeichen hierfür mehren sich zusehends (siehe auch die theoretischen Reflexionen Jägers) –, sollten echte Fortschritte in der „*Begabungs*“- wie „*Intelligenz*“-Forschung schon in allernächster Zeit möglich werden. Davon dürfte nicht zuletzt die praktische Intelligenz- bzw. Begabungsdiagnostik profitieren.

Spätestens seit Freud <sup>21</sup> wissen wir um die Bedeutung frühkindlicher Prägungseinflüsse für die spätere Persönlichkeitsentfaltung, insonderheit für die emotionale Entwicklung, aber auch die motivationale und Interessenbildung, das Arbeits- und Lernverhalten u.ä., kurz: im Hinblick auf den *Sozialisations*erfolg. Im sonderpädagogischen Kontext werden in diesem Zusammenhang die (möglichen) Folgen *sozialer* oder/und *sensorischer Deprivation* (vgl. u.a. Solomon et al. 1961) virulent. Deren Beziehung zur kognitiven, vorab sprachlichen, Entwicklung soll uns noch eingehend beschäftigen. Hier geht es zunächst weniger um das Inventar sozio-kultureller Einflußmöglichkeiten – darauf kommen wir im nächsten Abschnitt zu sprechen – als vielmehr um die grundsätzlichere Frage, ob bzw. in welchem *Ausmaß* sich die meßbare Intelligenz quantitativ im Laufe der Ontogenese verändert.

Daß die Beantwortung dieser Frage – neben ihrer pädagogischen Relevanz – für die *Intelligenzdiagnostik*, konkreter: im Hinblick auf die Zuverlässigkeit (Reliabilität oder Konsistenz) des Testurteils, von grundsätzlicher Bedeutung ist, wurde oben schon *angezeigt*. Nur unter der Voraussetzung relativ *konsistenter* Eigenschaften und Fähigkeiten, d.h. hier einer gewissen Zeitstabilität im Ausprägungsgrad meßbarer Intelligenz(dimensionen), lassen sich hinreichend genaue und verlässliche, als treffsichere, Intelligenzdiagnosen bzw. Begabungsprognosen i.S. erwarteter Lernleistungen stellen. Praktisch bedeutet dies, da sich beide Aspekte, der pädagogische Auftrag des Begabens wie das psychometrische Postulat nach relativ überdauernder Gültigkeit der Testurteile (Intelligenzdiagnosen), bis zu einem gewissen Grade wechselseitig bedingen, daß Intelligenzmessungen streng genommen nur je *aktuelle* (also mehr oder weniger kurzfristig gültige) Leistungsnachweise repräsentieren. Um Aussagen über das *potentielle* bzw. zukünftige Leistungsverhalten und damit Rückschlüsse auf das eigentliche Intelligenzpotential zu ermöglichen, bedarf es zusätzlich zur Anwendung traditioneller Intelligenztests eingehender Analysen des sozio-kulturellen Bedingungsgefüges der Intelligenz, also einer möglichst objektiven Erfassung *außerintellektueller* Faktoren wie Bildungsinteressen, Lernmotivation, familiärem Sprachverhalten und Kulturmilieu, Lehrerverhalten und anderer relevanter Schulvariablen, resp. der Abschätzung ihrer wahrscheinlichen Wirksamkeit in bezug auf die erwarteten Intelligenzleistungen des Schülers. M.a.W.: Der Begriff „Intelligenz“ mag zwar als abstraktes Konstruktum gedeutet werden, seine sichtbaren Wirkungen (die gemessenen Intelligenzleistungen) sind stets in konkrete (persönlich-motivationale und soziokulturelle) Bedingungsmuster eingebettet und von diesen mitbestimmt. Über den Grad der *Dynamik* dieser Determinanten intellektuellen Verhaltens (sub specie Aktualität versus Potentialität) mögen uns nun folgende Ausführungen genauer informieren.

#### a) Das Problem der Invarianz menschlicher Eigenschaften und Fähigkeiten

Einschlägige Längsschnittuntersuchungen – in den USA sind seit über 30 Jahren entsprechende Longitudinalstudien angesetzt<sup>22</sup> – belegen, daß die menschliche Intelligenz in der Ontogenese (z.T. beträchtlichen) Veränderungen unterworfen ist. Es scheint, daß die all-

---

<sup>21</sup> Wir können hier nur pauschal auf die reichhaltige tiefenpsychologische Literatur zum angeschnittenen Fragenkomplex verweisen. Für eine erste, rasche Problemorientierung empfiehlt sich Pongratz (1960).

<sup>22</sup> Es würde hier nur verwirren, die Fülle der Literaturquellen detailliert nachzuweisen. Der interessierte Leser sei diesbezüglich auf Bloom (1966) sowie die Zusammenfassung bei Lückert (1969b) verwiesen.

gemeine Intelligenz<sup>23</sup> in den ersten 4 bis 5 Lebensjahren am stärksten von solchen intraindividuellen Variationen betroffen ist und gegen Ende der Pubertät einen relativ hohen Stabilitätsgrad erreicht. Setzt man die „Endintelligenz“ auf 17 Jahre (Prozentuierungsbasis) fest, so ergibt sich nach Bloom (a.a.O., S. 64 ff.), daß bei normalem (durchschnittlichem) Verlaufstempo im Alter von 1 Jahr 20 %, von 4 Jahren 50 %, von 8 Jahren 80 % und von 13 Jahren 92 % der Intelligenzentwicklung abgeschlossen sind. Dies würde bedeuten, daß bereits im Alter von 4 – 5 Jahren die Hälfte der Endintelligenz bzw. mit 8 Jahren 80 % und mit 13 Jahren über 90 % der Erwachsenenintelligenz (Intelligenz-Optimum) vom normalsinnigen Kinde erreicht werden. Im Alter von 17/18 Jahren stabilisiert sich die Intelligenzentwicklung (die „Intelligenzkurve“ verläuft auf einem Plateau), bis sich im Alter wieder eine rückläufige Entwicklung anbahnt, deren Einsatz bzw. Verlaufsgeschwindigkeit nach den bisherigen Ermittlungen sehr großen individuellen Schwankungen unterliegt. Nach Wechsler beginnt der Intelligenzabbau – analog zum physiologischen Abbau – bereits zwischen dem 25. und 30. Lebensjahr, nach anderen Forschern erst im vierten oder fünften Lebensjahrzehnt, wobei jedoch Extremvarianten nach oben hin (besonders unter „Geistesarbeitern“) verhältnismäßig häufig beobachtet werden; von psychopathologischen Fällen abgesehen fällt die Abbaukurve der (allgemeinen) Intelligenz über viele Jahre oder Jahrzehnte hinweg zunächst ganz allmählich ab und ist erst im hohen Alter durch rapide Absenkungen charakterisiert.

Angeichts des ontogenetisch aufweisbaren (degressiven) Intelligenzzuwachses im Kindes- und Jugendalter – die weiteren Perioden interessieren im Zusammenhang unserer Gegenstandserörterung nicht – erhebt sich nun die überaus wichtige Frage, wie zuverlässig bzw. zeitstabil unter der Voraussetzung solcher Merkmalsvariation überhaupt Intelligenzbestimmungen möglich sind. Auch hierzu liegt eine Reihe von angloamerikanischen Untersuchungen vor. Demnach ist das Testergebnis (im Stanford-Binet, Wechsler-Bellevue oder einem ähnlichen Verfahren) vom 2. Lebensjahr (Lbj.) mit  $r = 0,41$ , vom 3. Lbj. mit  $r = 0,65$ , vom 4. Lbj. mit  $r = 0,71$ , vom 5. Lbj. mit  $r = 0,80$ , vom 8. Lbj. mit  $r = 0,90$  und vom 11. Lbj. mit  $r = 0,92$  mit der Endintelligenz der 17jährigen Probanden (immer im *intra-*

---

<sup>23</sup> Gewöhnlich wurde das Intelligenzniveau bzw. die Intelligenztest-Gesamtleistung den Altersvergleichen zugrunde gelegt. Die Hypothese eines g-Intelligenzfaktors ist freilich nach neueren faktorenanalytischen Befunden umstritten (vgl. S. 14ff. in diesem Buch). So verschieben sich in den von Hopf (1971) angestellten Berechnungen je nach verwandter Testmethode (z.B. Binet-Skala oder HAWIK) und/oder Basistheorie der Intelligenz die Zeitkurven kognitiver Entwicklung nicht unbeträchtlich. In bezug auf das multifaktorielle Modell der Primärfähigkeiten gelten etwa folgende Halbzeitwerte (für die in Klammern aufgeführten Dimensionen sensu Thurstone): 7;3 J. (P), 7;10 J. (R), 9;1 J. (S), 9;5 J. (N), 10;0 J. (M), 10;0 J. (V) und 13;3 J. (W). Und zu 80 % sind die genannten Intelligenzdimensionen erst auf folgenden Altersstufen entwickelt: 12 J. (P), 14 J. (R), 14 J. (S), 16 J. (N), 16 J. (M), 18 J. (V) sowie nach 20 J. (W). Im Hinblick auf die Möglichkeiten des Begabens im Schulalter ist deshalb pädagogischer Fatalismus m.E. fehl am Platze.

<sup>24</sup> Ein Korrelationskoeffizient in Höhe von 0.8 bedeutet eine optimale Übereinstimmung von rund zwei Dritteln aller Fälle, in Höhe von 0.9 jedoch bereits eine Übereinstimmungsquote von über 80 %. In der Regel ist man in der Testdiagnostik schon sehr zufrieden, wenn man Trefferquoten in dieser Höhe erzielt. Unaufgeklärte Restvarianzen von 10-20 % (Fehlprognosen) sind angesichts des multifaktoriellen Bedingungsgeflechts schulischer Lernleistungen vorerst unvermeidlich, wenngleich eine solche Feststellung keineswegs beruhigen, sondern eher die Forschungsinitiativen bekräftigen sollte. Bedenkt man jedoch andererseits, daß vergleichbare Lehrerurteile einen mittleren Voraussagewert von maximal 0.5 oder – gelegentlich – 0.6 (30-40 % Treffer) erreichen, dann zeigt sich erst deutlicher die Überlegenheit standardisierter Intelligenz-Leistungstests im Schul- und Bildungsbereich. Detailliertere Hinweise hierzu finden sich bei Chauncey und Dobbin (1968), Heller (1969b, 1970a), Ingenkamp (1969), Weingardt (1969) u.a.

individuellen Längsschnittmittel) korreliert. M.a.W.: Erst ab dem 7./8. Lbj. – unter Vorbehalt bzw. erhöhter Fehlerrisikobereitschaft schon ab dem 5. Lbj., jedoch kaum früher – werden einigermaßen verlässliche Intelligenzdiagnosen resp. Begabungs-(Lernleistungs-) Prognosen, wenigstens im Hinblick auf den Entwicklungsstand der 17-jährigen, möglich<sup>24</sup>. Siehe auch Bloom (1967).

Um die oben aufgeworfene Frage nach der Verlässlichkeit von Intelligenztestungen auf einer höheren Abstraktionsebene zu beantworten, könnte man jetzt auch sagen, daß die Aussagen von (allgemeinen) Intelligenztests umso zuverlässiger werden, je weiter in der Ontogenese fortschreitend sich das Merkmal „Intelligenz“ stabilisiert. Dabei unterstellen wir allerdings die Richtigkeit der g-Faktor-Hypothese sensu Spearman et al. Im Lichte multifaktorieller Intelligenzkonzepte (Thurstone, Jäger u.a.) müßte man die vorstehend referierten, empirisch-operational gewonnenen Befunde dahingehend interpretieren, daß offenbar in den verschiedenen Lebensstadien *nicht* die gleichen Fähigkeiten gemessen werden. Im Gegensatz zur Akzentuierung sinnesphysiologischer resp. motorischer Aufgaben in den Intelligenztestreihen für die jüngeren Probanden (Performance-Tests) findet sich in den Intelligenztests für ältere Schüler vorzugsweise kognitives bzw. sprachgebundenes, z.T. auch zahlengebundenes Aufgabenmaterial. Für die zweite Interpretationshypothese spricht die Erfahrungstatsache, daß im allgemeinen zwischen den Verbaltests und den psychomotorischen Tests sehr geringe, häufig dem Nullwert angenäherte Korrelationswerte gefunden wurden, was zumindest den Schluß zuläßt, daß mit den sog. Intelligenztests für die Altersstufen 1 bis 4 vorwiegend (psychomotorische) Entwicklungskomponenten und weniger Intelligenzfunktionen im engeren Sinne erfaßt werden. Eine letztgültige Entscheidung zugunsten der einen oder anderen Intelligenztheorie kann von hieraus freilich nicht getroffen werden; die empirischen Resultate sind mehrdeutig interpretierbar. Die beiden aufgewiesenen Tatbestände der *Intelligenzentwicklung* (Merkmalsvariation in der frühen Ontogenese) sowie der etwa ab dem Einschulungsalter *zunehmend verlässlicheren Intelligenzmessung* bleiben zunächst in ihrer Geltung von kontroversen Deuterversuchen der aufgezeigten Art unberührt. Für die Testdiagnostik und die schulpädagogische Praxis viel bedeutsamer erscheint in diesem Zusammenhang die Frage nach Art und Umfang derjenigen Bedingungskomponenten, die – außer den Anlagepotenzen – die Intelligenzentwicklung fördern oder hemmen können. Damit wäre erneut das Anlage-Umwelt-Problem thematisiert.

Die von Bloom (1966, S. 68 ff.) mitgeteilten Befunde aus Untersuchungen an eineiigen (EZ) und zweieiigen Zwillingen (ZZ) – vgl. als Pendant dazu den kritischen Beitrag von Ritter und Engel (1969) – lassen in guter Übereinstimmung zu den Ergebnissen von Gottschaldt (siehe unsere Ausführungen auf Seite 12 f.) erkennen, daß die Wirksamkeit der Erbanlagen im allgemeinen höher als die der Umwelteinflüsse zu bewerten ist. Allerdings divergieren die Quoten-Angaben der einzelnen Autoren beträchtlich (60 bis 90 % Anlage-Anteile). Immerhin konnten in einer Reihe von Vergleichsuntersuchungen, wobei einmal EZ-Paare, ein andermal ZZ-Paare unter sonst gleichen versus verschiedenen Milieubedingungen hinsichtlich ihrer Intelligenzentwicklung kontrolliert wurden, durch Umwelteinflüsse bewirkte positive bzw. negative IQ-Abweichungen bis zu 20, in Ausnahmefällen bis zu 30 Punkten beobachtet werden (Bloom, a.a.O., vgl. noch Anastasi 1958, Hunt 1961, Kirk 1962 u.a.). Die ersten 6 bis 8 Lebensjahre scheinen dabei besonders prägsam, d.h. für Bildungsanregungen jeglicher Art besonders aufnahmebereit versus verhängnisvoll (im Hinblick auf die spätere Intelligenzentwicklung) bei ungünstigen Milieuverhältnissen zu sein (vgl. soziokulturelle Deprivationshypothese). Jedenfalls sind Zusammenhänge zwischen sozio-kultureller Vernachlässigung und Retardierungserscheinungen auf der einen Seite sowie zwischen positivem (familialem resp. schulischem) Anregungsmilieu und Intelligenzzuwachs auf der anderen Seite unübersehbar. *Vorschulische Bildungsförderung* müßte sich demnach besonders wirksam bei sozio-kulturell benachteiligten Kindern erweisen, was durch Kirk (1958), Bloom et al. (1965), Nickel (1969) u.a. bestätigt werden konnte.

Andererseits gibt es nachweislich kritische Phasen in der intellektuellen Entwicklung, die bestimmte Lernvorgänge erleichtern oder erschweren, wie am Beispielfall der *Schulreife* (vgl. Aebli 1969, Kemmler & Heckhausen 1967, A. Kern 1970) eindrucksvoll deutlich wird. Siehe auch das Sammelreferat von Weinert (1964).

„Offensichtlich gibt es eine Zeit, in der der Mensch und das Tier bestimmte verpaßte Lernmöglichkeiten nicht mehr oder nur mit großer Mühe nachzuholen vermögen. Es scheint – mit anderen Worten – in der Geschichte der Organismen Perioden zu geben, in denen bestimmte Lernprozesse leichter als in anderen möglich sind. Theoretisch besonders wichtig sind jene Fälle, in denen ein Lernprozeß bis zu einem bestimmten Zeitpunkt *noch nicht* oder nur schwierig vollzogen werden kann (Phase der Unreife), worauf eine ‚kritische‘ oder ‚sensible‘ Periode folgt, in der der Lernprozeß – z.T. sehr leicht – gelingt, während in einer dritten Phase die Lernmöglichkeit wieder abnimmt. In vielen Fällen fehlt entweder die erste oder die dritte Phase“ (Aebli, a.a.O., S. 178).

Obwohl man die Reife-Komponente eher dem Anlagekomplex zuzuordnen geneigt ist, sind nach einigen erfolgreichen Versuchen, reifeabhängige Lernfunktionen – außerhalb der „sensiblen“ Phasen – zu trainieren (Klauer 1967, 1969; Nickel 1967, 1969 u.a.), auch hier Umwelteinflüsse nicht grundsätzlich von der Hand zu weisen. Es scheint, daß es so etwas wie eine absolute „strukturelle Leistungsobergrenze“ (Aebli) nicht gibt, zumindest ließ sich bis dato eine solche nicht belegen. Resümierend können wir somit nach Aebli „die *geistige Entwicklung* als Funktion von Anlage, Reifung, Umwelt- und Erziehungsbedingungen“ definieren. Ähnlich sieht Bloom (1966, S. 79 f.) in der Begabung (Bg) resp. Begabungsleistung (Intelligenz) ein Interaktionsprodukt, das er als Funktion von genetischen Potentialen (GP) und sozio-kulturellen Milieueinflüssen (SKM) beschreibt. Sofern man hierbei den Leistungsaspekt in den Blickpunkt rückt, müßte man diese Funktionsformel um einen dritten Faktor, die Persönlichkeits-Motivationsvariable (PM), erweitern <sup>25</sup>:

$$Bg = f (GP + SKM + PM)$$

Die Frage, ob Intelligenz eine fixe Größe sei, muß nach den bisher erörterten Informationen entschieden verneint werden. Weiterhin kann aus der Tatsache der Variierbarkeit von Intelligenzfunktionen geschlossen werden, daß Intelligenz nicht ausschließlich durch Anlage resp. Erbinformation festgelegt ist. Wäre dies nämlich der Fall, dürfte sie durch keinerlei Einflüsse (exogener Art) veränderbar sein. Eindeutig konnten jedoch milieuhängige intraindividuelle Merkmalsvariationen nachgewiesen werden. In Übereinstimmung mit den Befunden der Zwillingsforschung kann somit als gesichert festgehalten werden, daß Intelligenz eine *variable* Größe („Variable“ i.e.S.) darstellt, deren Spielraum innerhalb gewisser, durch Anlage und Vererbung keineswegs starr fixierter, Grenzen liegt. Dabei scheint freilich die – wenngleich im konkreten Falle selten exakt angebbare – obere intellektuelle Leistungsgrenze stärker anlageabhängig zu sein, während die peristatischen Determinationsanteile an der Gesamtvarianz der Intelligenz(leistung) durchschnittlich auf ein Drittel (im Verhältnis zur Anlagevarianz) geschätzt werden. Verständlicherweise kann es sich hierbei nur um sehr grobe Orientierungswerte handeln.

Die nächste Frage gilt jetzt natürlich den *Einzelfaktoren*, die die skizzierten Veränderungseffekte bewirken; nach Aufstellung obiger Funktionsformel beträfe dies die Faktorenbündel SKM und PM. Auch dazu liefert uns das Buch von Bloom aufschlußreiche Informationen, die durch die bekannten Arbeiten von Bernstein (1958, 1959, 1962, 1964, 1967) sowie das sehr lesenswerte Sammelreferat von Oevermann (1969) eine wertvolle Ergänzung erfahren. Siehe u.a. noch Lückert (1969b), Gaus (1969) und Spitzmüller (1969).

---

<sup>25</sup> Siehe dazu ausführlicher Heller 1969b, S. 380 f.

b) Sozio-kulturelle Determinanten intellektuellen Verhaltens (unter besonderer Berücksichtigung der Sprache)

Nach Bloom (1966, S. 76 ff.), der die Ergebnisse zahlreicher Forscher referiert, sind für die Intelligenzentwicklung vor allem folgende Umweltbedingungen förderlich: Ermutigung zu sprachlicher Aktivität des Kindes durch vorbildliche (sprachliche) Verhaltensmodelle, optimale Lernmöglichkeiten und Gelegenheiten zu ausgiebigen, intensiven Umwelterfahrungen, besonders auch Interaktionsmöglichkeiten mit Erwachsenen, leistungsmotivierende Milieueinflüsse, Aufgabenstellungen, die „klares Denken“ erfordern, Problemlösungen ermöglichen u. dgl. m. Wolf (1963) nennt insgesamt 13 Prozeßvariablen, die er drei Hauptgruppen zuordnet:

- I. *Stärke der Leistungsmotivation*
  1. Erwartung des Kindes bezüglich seiner geistigen Entwicklung
  2. Hoffnungen der Eltern bezüglich der geistigen Entwicklung des Kindes
  3. Interesse der Eltern an der geistigen Entwicklung des Kindes
  4. Art der Belohnung bei geistigen Fortschritten des Kindes
- II. *Förderung der Sprachentwicklung*
  5. Anhalten des Kindes zum Gebrauch der Sprache in verschiedenen Situationen
  6. Angebot an Gelegenheiten zur Wortschatzerweiterung
  7. Nachdruck auf korrekten Sprachgebrauch
  8. Qualität der gebotenen Sprachmodelle
- III. *Umfang des Lernangebots und der Lerngelegenheiten*
  9. Gelegenheiten zum Lernen innerhalb der Familie
  10. Lerngelegenheiten im Kindergarten und in der Vorschule
  11. Lerngelegenheiten außerhalb der Familie (und der Schule)
  12. Vorhandensein von Büchern (einschließlich von Nachschlagewerken)
  13. Beistand zur Erleichterung des Lernens in verschiedenen Situationen <sup>26</sup>.

Zwischen diesen Variablen und dem (allgemeinen) IQ fand Wolf eine multiple Korrelation in Höhe von  $r = 0.76$ , was enge Zusammenhangsverhältnisse andeutet. Im allgemeinen korrelieren nämlich die Milieuvvariablen – Sozialstatus der Kindesfamilie resp. Beruf und Ausbildungsstand der Eltern – kaum höher als 0.40 mit der Intelligenz.

Die meisten Begabungsforscher vertreten heute die Ansicht, daß die Wirksamkeit sozio-kultureller Einflüsse auf die intellektuelle Entwicklung in den Vor- und Grundschuljahren am stärksten sei. Dabei kommt naturgemäß in den ersten sechs Lebensjahren der *familiären* bzw. außerschulischen Umwelt die größte Bedeutung im Hinblick auf die kognitive Entwicklung des Kindes zu. In den Vordergrund der Betrachtung rücken hierbei a) der Aufbau der Leistungsmotivation, b) die Formen der Kindererziehung und c) die besonders von Bernstein herausgearbeiteten schichtspezifischen Sprachcodes.

#### *Aufbau der Leistungsmotivation*

Die Leistungsmotivation oder das Leistungsstreben stehen ohne Frage in direkter Beziehung zum Schul- und Studienerfolg (resp zur Berufsleistung). Ihr Einfluß auf die Entwicklung der Intelligenz und Begabung, zumal im Vorschulalter, tritt dagegen zunächst weniger deutlich in Erscheinung. Und trotzdem ist die frühkindliche Erziehung zum Leistungsstreben gerade auch im Hinblick auf die intellektuelle Förderung ungem. wichtig. Nach den Untersuchungsergebnissen von McClelland (1955, 1961) und seinem Mitarbeiterkreis – siehe auch das Sammelwerk von Atkinson (1958) – unterscheiden sich die Eltern hochleistungsmotivierter Kinder sehr deutlich von denen niedrig motivierter unter dem Gesichtspunkt frühkindlichen Erziehungsverhaltens.

---

<sup>26</sup> Zit. nach der Übersetzung von H.R. Lückert 1969 b, S. 155.

So fand z.B. Winterbottom (zitiert nach Atkinson 1958, S. 433 ff.; vgl. auch Rosen 1956, McClelland et al. 1953, McClelland loc. cit.), daß die Mütter hochleistungsmotivierter Söhne früher und nachhaltiger Wert auf Selbständigkeit (im Handeln, in Entscheidungssituationen etc.), Unabhängigkeit von Erwachsenen, Selbstkontrolle, Aktivität und Individualität ihrer Kinder legten als die Mütter niedrigleistungsmotivierter Söhne. Diese verlangten allenfalls in „äußeren“ Dingen (z.B. zur eigenen Entlastung im Haushalt) ihren Kindern mehr ab, legten ihnen jedoch sonst weit mehr Freiheitsbeschränkungen auf und erwarteten erst viel später Unabhängigkeit und Selbständigkeit in persönlichen Entscheidungssituationen. Die Mütter hochleistungsmotivierter Kinder waren dagegen zunächst darauf bedacht, ihren Kindern Selbstvertrauen zu geben und kamen mit zahlenmäßig bedeutend weniger Restriktionen aus. Diese setzten sie vor allem zur Abwehr regressiver Verhaltensweisen ein, ermutigten aber gleichzeitig ihre Kinder (Söhne) zur Meisterung sozialer und kognitiver Probleme, deren Erfolge sie positiv bekräftigten.

Nach Kagan et al. (1958) korrelieren nun diese Eigenschaften versus Erziehungsmerkmale positiv mit dem IQ-Zuwachs. D.h. Selbständigkeit im Verhalten, größere emotionale Unabhängigkeit, positive Einstellung zu Problemlösungen, Freude an Wettbewerben, Aggressivität, Aufschiebbarkeit unmittelbarer Triebbedürfnisse u.ä. wie überhaupt eine betont zukunftsorientierte Einstellungshaltung (größere „Zeitperspektive“) müssen demnach als intelligenz- und begabungsfördernde Faktoren motivationaler bzw. sozialkultureller Art betrachtet werden. Daß diese Merkmale – ausgenommen die Aggressivität – typische Mittelklasse-Erziehungsideale repräsentieren, sei mit Spitzmüller nur am Rande erwähnt. Wichtiger ist hier der Hinweis McClellands, daß bereits um das 6. Lebensjahr ein erstes Optimum-Alter im Aufbau der Leistungsmotivation durch „mütterlichen Leistungsdruck“ beginnt. Überhaupt scheint zu Beginn der motivationalen (und kognitiven) Entwicklung der Mutter die weitaus bedeutsamere Erziehungsrolle zuzufallen, dies kann zumindest im Hinblick auf die Söhne aufgrund vorliegender Untersuchungsergebnisse behauptet werden (siehe dazu noch Heckhausen 1966 u. 1969, Heller 1969 b u. 1970 a sowie neuerdings Keil und Keil-Specht 1970). Eine „fordernd-verstehende“ oder „verstehend-fordernde“ (sensu Ruppert) Erziehungshaltung ist demnach dem Streben nach intellektueller Tüchtigkeit in diesem frühen Alter am förderlichsten.

### *Formen der Kindererziehung*

Es wurde schon mehrfach – zuletzt im Zusammenhang mit der Erörterung der Motivationsgenese und ihrer Bedeutung für die kognitive Entwicklung – angedeutet, daß die Methoden der Kindererziehung die Begabungsentfaltung fördern oder hemmen können. Andererseits scheinen die jeweils praktizierten Erziehungsstile hinwiederum abhängige Variablen der sozialen Schichtverhältnisse darzustellen. Nach Bronfenbrenner (1958), der eingehende Merkmalsanalysen hierzu anstellte, sind für die schichtspezifischen Einstellungen und Verhaltensweisen im Kontext der Erziehung vor allem folgende Züge charakteristisch (zit. nach Spitzmüller 1969 S. 190): „Einerseits ist die *Mittelschicht* nachgiebiger gegenüber ihren Kindern, läßt ihnen mehr Freiheit, gibt ihnen zu essen, wenn sie hungrig sind, ist toleranter in der Reinlichkeitsdressur und in der Sexualerziehung, andererseits erwartet sie auch mehr von ihnen: sie sollen früher selbständig werden und vor allem sollen sie in der Schule möglichst weit kommen, also auf das College gehen. Die *Unterschicht* hingegen ist weniger nachgiebig, verlangt aber auch weniger Selbständigkeit. Wie nun erreicht die Mittelschicht ihre höheren Erwartungen, da sie doch zugleich nachgiebiger ist?

Diese Frage wird beantwortet, wenn man sich die Methode der Erziehung näher anschaut. Während die Unterschicht sich eher auf die körperliche Züchtigung als Erziehungsmittel verläßt, arbeitet die Mittelschicht mehr mit Argumenten, Iso-

lierung, Appellen an das Gewissen und anderen Methoden, die den Verlust der elterlichen Liebe implizieren. Da die Mutter aber in der früheren Kindheit liebevoller und nachgiebiger war, wirkt diese Methode besonders gut, denn der Verlust wird umso schmerzlicher empfunden, je enger die Beziehungen waren. Aggressivität andererseits, wie sie in der körperlichen Züchtigung zum Ausdruck kommt, vermehrt in dem Bestraften eher seine Aggressionen, bestärkt also das unerwünschte Verhalten anstatt es zu vermindern". M.a.W.: Während die Mütter der Unterschicht sehr stark Gesichtspunkte äußerer Ordnung und Wohlverhaltens (Regelmäßigkeit, Sauberkeit, Pünktlichkeit u.ä.) sowie Unterordnung resp. Gehorsam in ihrer Kindererziehung betonen, legen die Mütter der Mittelschicht mehr Wert auf verstehende Führung (Verständnisweckung, Argumentation mit Vernunftgründen), emotionale Bindung und Wärme (liebevolle Beziehungen, Wunsch und Glück und Zufriedenheit ihrer Kinder), Gewissensappelle (Drohung mit Liebesentzug, Isolierung etc.), Selbstkontrolle des Kindes, Weckung der Lernmotivierung, Bildungs- und Berufsansprüche u. dgl. m. Nach Hieronymus hängt der soziale Status sogar enger mit dem Niveau der sozio-ökonomischen Erwartung zusammen als mit der Intelligenzleistung. Von hieraus wird auch die bereits oben erwähnte größere Zeitperspektive von Mittelschichtkindern (und Eltern), z.B. bezüglich langfristiger Ausbildungsziele, verständlich.

Bernstein (1967, S. 352 ff.), der weitere einschlägige Untersuchungen referiert, hält in diesem Zusammenhang den Unterschied zwischen „Absicht“ und „Folgen“ im Hinblick auf den Sozialisationserfolg für aufschlußreich. „Zwei wichtige Untersuchungen von Kohn (1959) über das Verhältnis zwischen verschiedenen Arten von Disziplin und sozialem Status weisen darauf hin, daß Eltern aus der Arbeiterklasse eher im Hinblick auf die *unmittelbaren Folgen* der kindlichen Tat reagieren, während Eltern aus der Mittelklasse eher im Rahmen ihrer Interpretation der *Absicht* reagieren, die das Kind im Augenblick seiner Tat verfolgt hat. Sowohl die Mittel- als auch die Arbeiterklasse handeln im Hinblick auf entfernte Ziele; diese Ziele sind jedoch verschieden. In der Arbeiterklasse gruppieren sich die Werte der Eltern rund um den Begriff der Achtbarkeit oder Respektabilität (vgl. Duvall), während sie sich in der Mittelklasse auf verinnerlichte Standards des Verhaltens konzentrieren. So handeln die Eltern aus der Arbeiterklasse im Hinblick auf Ziele, die darauf gerichtet sind, ungehorsame und schimpfliche Taten zu verhindern, während die Eltern aus der Mittelklasse auf Absicht und Tat unter Bezugnahme auf individualisierte Standards reagieren. Die Eltern aus der Arbeiterklasse nehmen keine oder nur in einem minimalen Maß Rücksicht auf die subjektive Absicht. Die Bedeutung, die diesem Unterschied für die spätere Entwicklung des Kindes zukommt, kann nicht genügend hervorgehoben werden.“

Die skizzierten Differenzen schichtspezifischer Erziehungsstile, wie sie vor allem in der Nachkriegszeit erforscht wurden, dürfen freilich nicht (ideologisierend) als unverrückbare Maximen interpretiert werden. So konnte beispielsweise Bronfenbrenner (loc. cit.: vgl. auch Herrmann 1966, bes. S. 63 ff.) nachweisen, daß sich im Laufe der letzten 30 Jahre in der amerikanischen Bevölkerung ein deutlicher Stilwandel vollzogen hat. Demnach zeigten sich in der Zeit vor dem Zweiten Weltkrieg die der Mittelschicht zugerechneten Eltern strenger und wesentlich restriktiver (z.B. hinsichtlich Fütterungsgewohnheiten, Abstillen, Sauberkeitserziehung) oder – einer Modebezeichnung folgend – autoritärer, wohingegen sich im gleichen Zeitraum die Mütter der Arbeiterklasse allgemein offenbar großzügiger und liberaler gaben (sie bevorzugten Brustfütterung, stillten auf Verlangen des Kindes, Abstillen und Sauberkeitserziehung erfolgten viel später etc.). Der Umschwung im letzten Weltkrieg wird meistens mit dem Hinweis auf neuere Forschungsergebnisse der Psychologie, Pädagogik und Medizin bzw. die besseren Informationsmöglichkeiten der Mittelschichtbevölkerung (Zugang zu Büchern, Zeitschriften u.ä.) erklärt. Sofern man hier die Wirksamkeit der modernen Kommunikationsmedien (Radio, Fernsehen, Presse) gerade auch für die Angehörigen der sog. Unterschicht in Rechnung stellt, müßte in



nicht allzuferner Zeit – die Richtigkeit obiger Erklärungshypothese unterstellt – ein erneuter Umschlag bzw. Ausgleich in Einstellung und Erziehungspraktiken zwischen den Sozialschichten zu erwarten sein. Dies ist zumindest plausibel und denkbar. Im Hinblick auf die Zusammenhänge zwischen Erziehungsstil und Bildungserwartung versus Begabungsentwicklung wäre ein solcher Trend zu sozialem Ausgleich über berechnete sozialpolitische Forderungen hinaus sogar höchst wünschenswert. Freilich liegen die diesbezüglichen Probleme sozialer Determination kognitiver Vollzüge und ihrer Differenzen viel tiefer, wie gleich noch zu zeigen sein wird. Immerhin konnte u.a. Floud (1956) aufgrund umfangreicher empirischer Forschungen im Rahmen der Sekundarschulbildung nachweisen, daß beispielsweise Arbeiterkinder einen unverhältnismäßig höheren IQ aufweisen müssen, um mit ihren Schulkameraden aus der Mittelschicht auf der Grammar school leistungsmäßig Schritt halten zu können. Unsere allerjüngsten Erfahrungen in der baden-württembergischen Bildungsberatung können diese Ergebnisse unschwer bestätigen (siehe u.a. Heller 1969 b, 1970), so daß die Bedeutung des Erziehungsmilieus für die Begabungsentfaltung auch hier evident wird. Ein sehr wahrscheinlich noch gewichtigerer Grund für die soziale Disproportion von Begabungsausprägung resp. aktuellen Schulleistungsverhältnissen liegt freilich – jenseits vielberufener systemimmanenter Ursachen – im schichtenabhängigen Sprachverhalten, womit wir endlich zum nächsten Punkt überleiten.

### *Schichtspezifische Sprachkodes*

Der Einfluß des Sprachverhaltens auf die kognitive Entwicklung wurde wohl nie ernsthaft in Frage gestellt. Trotzdem hat die eigentliche (empirische) psycholinguistische Forschung erst vor wenigen Dezennien begonnen (vgl. Hörmann 1967). In diesem Kontext gewinnt die Bernsteinsche Theorie des Sprachgebrauchs – deren Kern die Interdependenz von Sprache und Welterfahrung betont (d.h. die Art und Weise, wie wir denkend und erlebend die *Umwelt erfahren*, hängt wesentlich von der durch die jeweilige Sozietät geprägten *Sprache* ab) – im Hinblick auf die Begabungsentwicklung hervorragende Bedeutung. Die hier von Bernstein (1967, S. 346) vertretene These lautet: „Die Implikationen bestimmter Formen von Sprachgebrauch, wie er in einer normalen Bevölkerung zu finden ist, verzögern oder erleichtern das Lernen, und zwar unabhängig von jeder angeborenen Fähigkeit. Gewisse sprachliche Formen bringen für den Sprechenden einen Verlust an Geschicklichkeiten (sowohl kognitiven wie auch sozialen) mit sich, die sowohl für eine erfolgreiche Erziehung als auch für den Erfolg im Beruf von ausschlaggebender Wichtigkeit sind; diese Formen des Sprachgebrauchs sind kulturell und nicht individuell bestimmt.“ Diese von Bernstein (1958, 1962, 1967) analysierten Formen des Sprachgebrauchs werden nun bestimmten Sozialgruppen zugeordnet und hieraus begabungspsychologische sowie bildungssoziale Folgerungen abgeleitet. So wird heute vielfach die sprachliche (und intellektuelle) Retardierung der Kinder aus unterprivilegierten Sozialschichten (etwa der auf Gymnasium und Realschule unterrepräsentierten Arbeiter und Bauern) nicht mehr genetisch, sondern sozio-kulturell bedingt, genauer: die schichtspezifischen Formen des Sprachverhaltens für difizitäre Bildungsergebnisse verantwortlich erklärt. Hierbei sind nach Bernstein (1959, 1964) folgende Qualitätsunterschiede des *elaborierten* (formalen) versus *restringierten* (öffentlichen) Sprachkodes aufweisbar. „Die Merkmale der sprachlichen Form, an die das Kind aus der *Mittelklasse* gewöhnt wird, sind die folgenden:

1. Die logische Modifikation und der logische Akzent werden durch eine grammatisch komplexe Satzkonstruktion vermittelt, vor allem durch die Verwendung von Konjunktionen und Nebensätzen.
2. Häufiger Gebrauch von Präpositionen, die sowohl rein logische Beziehungen als auch zeitliche oder räumliche Nähe anzeigen.

3. Häufige Verwendung der unpersönlichen Pronomen „es“ und „man“.
4. Diskriminierende Auswahl aus einer Reihe von Adjektiven und Adverbien.
5. Der expressive Symbolismus, der durch diese sprachliche Form bedingt wird, verleiht dem Gesagten weniger logische Bedeutung als affektive Unterstützung.
6. Die individuelle Qualifikation wird verbal durch die Struktur und die Beziehungen innerhalb und zwischen den Sätzen vermittelt. Die subjektive Absicht wird unter Umständen in Worten erläutert.
7. Es handelt sich um eine Form des Sprachgebrauchs, die auf die Möglichkeiten hinweist, die einer komplexen begrifflichen Hierarchie inhärent sind und die die Organisation der Erfahrung erlauben.

Wir wollen diesen Typus des Sprachgebrauchs *formal* nennen.

In der *Arbeiterklasse* gehört die sprachliche Beziehung zwischen Mutter und Kind einer anderen Ordnung an. Es handelt sich im wesentlichen um eine sprachliche Form, bei der die persönlichen Qualifikationen zunächst mit Hilfe des expressiven Symbolismus ausgedrückt werden, das heißt in nichtsprachlicher Form oder mit Hilfe einer begrenzenden Sprachstruktur. Die Beziehung des Kindes zur Mutter ist unmittelbar. Sein wichtigster Bezugspunkt ist nicht ihre Sprache, denn die persönliche Qualifikation, das „Ich“ der Mutter aus der Arbeiterklasse kommt namentlich in einem expressiven Symbolismus zum Ausdruck, der ausschließlich auf sich selbst Bezug hat. Kurz, die subjektive Absicht wird nicht in Worten explizit gemacht oder erläutert. Das Kind lernt früh zu reagieren, und zwar auf Anhaltspunkte, die unmittelbar relevant sind. Es handelt sich also um eine Form der Kommunikation, in der die unmittelbare Erfahrung der affektiven Zusammengehörigkeit maximal hervorgehoben wird, wogegen die sprachlich bedingte emotionale und kognitive Differenzierung in den Hintergrund tritt. Das Kind aus der *Arbeiterklasse* ist daher für einen Sprachgebrauch empfänglich, der sich völlig von dem der Mittelklasse unterscheidet. Die Merkmale dieses Sprachgebrauchs sind die folgenden:

1. Kurze, grammatisch einfache, oft unfertige Sätze von dürftiger Syntax, die meist in der Aktivform stehen.
2. Verwendung einfacher und immer derselben Konjunktionen (so, dann, und).
3. Häufige Verwendung kurzer Befehle und Fragen.
4. Seltener Gebrauch der unpersönlichen Pronomen „es“ und „man“.
5. Starre und begrenzte Verwendung von Adjektiven und Adverbien.
6. Die Feststellung einer Tatsache wird oft im Sinne einer Begründung und einer Schlußfolgerung verwendet, genauer gesagt, Begründung und Folgerung werden durcheinandergeworfen, und am Ende entsteht eine kategorische Feststellung, wie „Du gehst mir nicht aus dem Hause“; „Laß das in Ruhe“.
7. Die individuelle Auswahl aus einer Reihe traditioneller Wendungen oder Aphorismen spielt eine große Rolle.
8. Feststellungen werden als implizite Fragen formuliert, die dann eine Art Kreisgespräch auslösen, bei dem sich die Gesprächspartner ihrer gegenseitigen Sympathie versichern, das heißt, man redet in gegenseitiger Übereinstimmung im Kreis herum. Zum Beispiel: „Stell dir vor?“; „Das hätte ich nicht gedacht?“. Oder: „Na, was sagen Sie von der Lotte?“ – Ist ein schlimmes Unglück. – Haben Sie recht. Ist wirklich ein schlimmes Unglück. Und die *armen Eltern*. – Ja, die *armen Eltern*. Was dachte sich das Mädchen dabei? – Dachte sich gar nichts, tat eben, was sie wollte. – Ein Unglück nenne ich das, ein schlimmes Unglück für das ganze Haus.“
9. Der Symbolismus besitzt einen niedrigen Grad der Allgemeinheit.
10. Die persönliche Qualifikation wird aus der Satzstruktur weggelassen oder ist nur implizit vorhanden, folglich wird die subjektive Absicht nicht mit Worten explizit gemacht oder erläutert.

Die Form des Sprachgebrauchs, welche durch diese Merkmale gekennzeichnet ist, wollen wir *öffentlich* nennen... Denn eine *öffentliche Sprache* ist in erster Linie ein Vehikel für die Schaffung von *sozialen*, nicht von *individuierten* Symbolen. Im Gegen-

satz zum Individuum, das einen Sprachgebrauch zu schaffen lernt, worin es auswählen kann, um seine persönlichen Gefühle zu vermitteln (wie bei einer *formalen Sprache*), neigt die Person, die sich einer *öffentlichen Sprache* bedient, dazu, ihre Gefühle an gesellschaftliche Redensarten und Etiketten zu hängen, welche die soziale Verbundenheit, die Solidarität maximal betonen, und zwar auf Kosten der logischen Struktur der Kommunikation und der Spezifität des Gefühls" (Bernstein 1967, S. 358 ff.).

Wie wirkt sich nun der Einfluß dieser schichtspezifischen Funktionsmuster der Sprache auf die kognitive Entwicklung aus? Allgemein fällt bei Begabungsuntersuchungen immer wieder auf, daß Unterschichtkinder in den *verbalen* Testleistungsdimensionen (im Gruppendurchschnitt) schlechter abschneiden im Vergleich zu den Kindern aus der Mittelschicht. Hingegen treten solche Unterschiede zwischen den Sozialklassen bei sog. nonverbalen Intelligenztestreihen weniger deutlich oder überhaupt nicht in Erscheinung. Ferner korrelieren die Variablen der Sprachtests höher mit dem Sozialstatus (Vaterberuf bzw. Schulbildung der Eltern) als die Variablen nichtsprachlicher Intelligenztestaufgaben, wobei sich diese Tendenz offensichtlich noch mit zunehmendem Lebensalter verstärkt. Dagegen scheint eine negative Korrelation zwischen der Höhe verbaler Intelligenzfunktionen einerseits und dem Ausmaß „sozio-kultureller Deprivation“ andererseits vorzuliegen (vgl. Oevermann 1969, S. 310). Alle genannten Ergebnisse deuten einen engen Zusammenhang von Sprache und Intelligenz-entwicklung versus Sprache/Intelligenzfunktion und sozialkulturellen Milieustrukturen an.

Oevermann (loc. cit., S. 311 ff.), der in seinem Sammelreferat die Untersuchungen zahlreicher Forscher auswertet, hebt in diesem Zusammenhang noch folgende Ergebnisse hervor. Die Lautdiskriminier- bzw. Artikulationsfähigkeit, die in den ersten Lebensjahren deutliche schichtspezifische Unterschiede (zugunsten der Mittelklasse) aufweist, korreliert mit der späteren Intelligenzleistung zwischen  $r = 0.24$  und  $0.71$ , wobei besonders wiederum Zusammenhänge zu den verbalen Intelligenzfunktionen (Wortschatz, syntaktische Komplexität, Lese-Lern-Erfolg) auffallen. Fast noch bedeutsamer als die quantitativen Leistungsdifferenzen erweisen sich *qualitative* Unterschiede beider Sprachkodes im Hinblick auf die Denk- und Intelligenz-entwicklung resp. Begabungsprägung. So steht etwa die Diversifikation des aktiven Wortschatzes, gemessen in TTR-Maßeinheiten<sup>27</sup>, in engerer Beziehung zur Testintelligenz als der absolute Umfang des Wortschatzes. Bei entsprechenden sprachlichen Produktionen waren die Unterschichtkinder den Kindern aus der Mittelklasse deutlich unterlegen. Dies besagt, daß sich die psycholinguistischen Codes (der beiden Sozialschichten) auf der lexikalischen Ebene stärker in der Art heterogener Wortwahl als im Gesamtindex produzierter Wörter unterscheiden.

„Wichtiger als der quantitative Umfang des Wortschatzes scheint uns dessen qualitative, *semantische Struktur* zu sein, Cohn (1955) und Riessman (1962) machen auf den Reichtum expressiver und konkret-deskriptiver Bezeichnungen in der Sprachkultur der Unterschicht aufmerksam und rühmen das Repertoire von fein und detailliert abgestuften Wörtern für die typischen Gegenstandsbereiche in dieser Schicht, ohne das allerdings durch Daten genauer belegen zu können. Diese Frage ist unseres Wissens bisher leider nicht systematisch untersucht worden; wenn man aber an die Eigenarten bestimmter Berufssprachen oder an die Vielfältigkeit der Wörter für „Geld“, „Polizei“ usw. in der Unterschicht denkt, erscheint die obige Annahme durchaus plausibel. Sie sollte allerdings auch nicht strapaziert werden. So zeigte sich, daß Kinder der Mittelschicht lexikalische Lücken durchaus kontextadäquat in einer für die Unterschicht typischen Sprach-

---

<sup>27</sup> Der *Diversifikations-Quotient* (TTR = Type-Token-Ratio) drückt das Verhältnis der verschiedenen Einzelwörter zur Gesamtzahl der Wörter einer Sprachprobe aus.

probe füllen konnten, während das umgekehrt für Kinder der Unterschicht für einen typischen Mittelschichttext nicht gilt (Deutsch u.a. 1964)”. (Oevermann, loc. cit.).

Man geht wohl nicht fehl in der Annahme, daß der mehr beschreibende (weniger analysierende) Stil des restringierten Sprachkodes in seiner Auswirkung auf die kognitiven Funktionen vorab die Begriffsbildung und die Entdeckung kausaler Ordnungszusammenhänge erschwert. Einige Autoren weisen in diesem Zusammenhang auch auf die Diskrepanz zwischen dem Wortschatz der Unterschichtkinder und dem Schul-(Lesebuch-)Wortschatz hin; diesbezügliche amerikanische Stichproben erbrachten selten Übereinstimmungsquoten, die über 50 % lagen. Auch hinsichtlich der *Wortarten* zeigen sich im Sprachverhalten sozialrelevante Unterschiede. „Auf verschiedenen Altersstufen gebrauchen Kinder der Mittelschicht mündlich und schriftlich bei kontrollierter Intelligenz in differenzierterer Form Adjektive, Adverbien und ungewöhnliche Präpositionen und Konjunktionen (Bernstein 1962, Lawton 1964, 1968, Oevermann 1967)... Jene Wortarten (Adverbien, Präpositionen, Konjunktionen), die in der Sprachentwicklung verhältnismäßig spät auftauchen, kommen in der Sprache der Unterschichtkinder weniger häufig vor. Dagegen ist der Anteil der Substantive in der Unterschichtsprache hoch (McCarthy 1930, Day 1932, Templin 1957, Thomas 1962, Oevermann 1967). McCarthy deutet das als Anzeichen einer wenig differenzierten Ausdrucksweise ” (Oevermann, loc. cit.).

Vielleicht die gravierendsten Folgen im Hinblick auf die Möglichkeiten intellektueller Daseinsbewältigung sind jedoch durch Kodeunterschiede in der *syntaktischen* Elaboration zu erwarten. Oevermann, der diesbezügliche Hypothesen durch eine Reihe experimenteller Einzelbefunde recht plausibel belegen kann, präzisiert und modifiziert andererseits zugleich einige Grundannahmen der Bernsteinschen Theorie (siehe hierzu ausführlicher die kritische Analyse von Niepold 1971). Im Zusammenhang von Sprache und Begabung interessieren nun besonders folgende psycholinguistischen Forschungsergebnisse der letzten Jahre (nach Oevermann 1969, S. 314 ff.).

„Die Analyse von Wortarten ist unter dem Gesichtspunkt der syntaktischen Organisation von größerer Bedeutung (R.W. Brown 1957). In Wort-Assoziations-Analysen wird zwischen Assoziationen in der identischen grammatischen Wortklasse (paradigmatische oder homogene Assoziationen) und in formal nicht-identischen Wortklassen (syntagmatische oder heterogene Assoziationen) unterschieden. Paradigmatische Assoziationen gelten als Zeichen für die Beherrschung der grammatischen Regel der Wortartenidentifizierung, während syntagmatische Assoziationen das Befolgen typischer assoziativer Verknüpfungen von Wörtern in Sätzen erkennen lassen. Dieser ‘primitiveren’ Assoziationsstruktur liegen vermutlich die Übergangswahrscheinlichkeiten zwischen Wörtern in der Umgangssprache zugrunde, während paradigmatische Assoziationen das Prinzip der Klassifikation in ein grammatisches Schema, die Anwendung einer abstrakten grammatischen Regel gegenüber einfachem Behalten implizieren. Diese Interpretation wird durch Experimente gestützt (Brown/Berko 1966, McNeill 1966). – So verwundert es nicht, wenn eine mit dem Alter zunehmende Tendenz zu paradigmatischen Assoziationen festzustellen ist (Entwisle 1966) und geistig retardierte Kinder weniger paradigmatische Assoziationen produzieren als normale Kinder (Simmel 1966). Französische Studenten assoziierten wesentlich häufiger im paradigmatischen Stil als französische Arbeiter (Rosenzweig 1964)... Insgesamt deuten die Ergebnisse an, daß bei Angehörigen der Unterschicht der Worteinfall weniger durch *abstrakte*, grammatische Regeln gesteuert wird.

Kann die grammatische Analyse von Assoziationen Hinweise über grammatische Regeln bei der Planung von sprachlichen Äußerungen geben, so erlaubt die Analyse der *syntaktischen Struktur* zusammenhängender Sprachsequenzen direkte Schlüsse auf die Sprachgewohnheiten und den sprachlichen Entwicklungsstand. Dazu liegt eine Reihe von Untersuchungen vor, die konsistent die ‘primitivere’ syntaktische Organisation in der Sprache der Kinder aus der Unterschicht belegen. Am häufigsten

ist die durchschnittliche Satzlänge gemessen worden. Zwischen dem zweiten und vierten Lebensjahr ist das rascheste Anwachsen auf diesem Maß zu beobachten. Die Sprache der Unterschichtkinder besteht aus durchschnittlich kürzeren Sätzen (Davis 1937, Day 1932, Templin 1957, Thomas 1962), auch wenn das Intelligenzalter oder der IQ und die Schulklassenzugehörigkeit kontrolliert werden (McCarthy 1930, Oevermann 1967). Die Unterschiede werden mit zunehmendem Lebensalter eher größer, auf keinen Fall findet aber eine Nivellierung der Differenzen statt...

Übereinstimmend ist von vielen Autoren der *Subordinationsindex* (zumeist der Quotient aus den abhängigen Nebensätzen und der Anzahl der finiten Verben) zur Messung der syntaktischen Komplexität als der beste Einzelindex für die Sprachentwicklung nach dem dritten Lebensjahr empfohlen worden. Die relevanten Untersuchungen zeigen hier sehr deutliche Unterschiede zwischen der Sprache der Kinder in der Mittel- und Unterschicht (Templin 1957, Thomas 1962, Loban 1964, Deutsch u.a. 1964)... Die Ergebnisse von Loban und Lawton zeigen übereinstimmend, daß sich die Schichtengruppen vor allem im Gebrauch adverbialer Nebensätze unterscheiden, während Oevermann (1967) den größten Unterschied für Relativsätze fand, wozu es eine Parallele im Vergleich der Schriftsprache von gehörlosen und normalen Kindern gibt (Heider/Heider 1940, vgl. auch die Zunahme von Relativsätzen in der Sprache älterer Kinder bei Templin 1957). – Loban/Lawton (1963) fanden in einem syntaktischen „multiple choice“-Test eine größere Sicherheit im Gebrauch sinnadäquater Konjunktionen bei den Kindern der Mittelschicht, und Oevermann (1967) stellte für die Sprache dieser Schicht einen häufigeren Gebrauch kausaler, finaler und konzessiver Konjunktionen fest.

Loban (1964) unterzog die syntaktischen Satzbaupläne einer detaillierteren Analyse und stellte fest, daß sich die Schichtengruppen weniger in den syntaktischen Grundstrukturen als in der Flexibilität der Erweiterung dieser Strukturen unterschieden. Das weist zusätzlich auf eine größere Rigidität der Satzplanung in der Sprache der Unterschicht hin, in der ein abwägendes Urteil, die gedankliche Konstruktion logischer Möglichkeiten und die Präzisierung von Information nur wenig durch die Verwendung etwa von Konzessiv- und Konditionalsätzen zum Ausdruck kommt.

Nach diesen Ergebnissen, die insgesamt eine syntaktisch wenig laborierte Struktur als vorherrschendes Kennzeichen der Unterschichtsprache verdeutlichen, erstaunt es nicht, wenn darin gleichzeitig eine größere Anzahl grammatischer Fehler gefunden wird (McCarthy 1930, Templin 1957, Thomas 1962, Loban 1964). Besonders häufig stimmen Numerus und Person zwischen Prädikat und Subjekt nicht überein. Außerdem finden sich häufig idiomatisch bedingte „Fehler“. Auslassungen des Hilfsverbs bei Konjugationen, falsche Verbformen und Fehler im Gebrauch der doppelten Negation (Thomas 1962, Loban 1964)<sup>28</sup>.

Mit diesen letzten Erörterungen berührten wir bereits Probleme der *Sprachentwicklung* und ihrer Fehlbildungen, die wir andernorts ausführlicher behandelten (Heller 1972a). Im Kontext sprachlicher Einflüsse auf die Kognition, insonderheit unter Berücksichtigung der Begabung, interessieren jetzt natürlich die Zusammenhänge zwischen den einzelnen linguistischen Indikatoren und der Intelligenzentwicklung bzw. dem Schulleistungsverhalten überhaupt. Auch dazu liegen empirische Untersuchungsergebnisse vor, wenngleich hier die spekulativen Beiträge aus der (philosophischen) Sprachpsychologie bislang überwiegen. So wird vielfach die Sprache als die bedeutendste Determinante kognitiver Entwicklungsprozesse angesehen. Sehr oft herrscht die Meinung vor, daß mit unseren sprachlichen Ausdrucksmitteln erst eigentlich die Denkprozesse konstituiert werden, wie überhaupt Sprache und Denken schon immer in enger Abhängigkeit gesehen wurden. Durch die menschliche Sprache werden die diffusen, ungliederten psychischen Prozesse weitgehend erst geordnet und strukturiert (vgl. u.a. Bereiter, Bernstein, C.P. Deutsch, M. Deutsch, Jensen). „Der größte Nachteil, der mit der Verwendung einer *öffentlichen* Sprache verknüpft

---

<sup>28</sup> Die Literaturangaben sind – soweit nicht im Index dieses Buches aufgeführt – vollständig im Nachweisregister Oevermanns (1969) einzusehen.

ist, besteht gerade in der Verständnislosigkeit gegenüber komplexen Verallgemeinerungen und gegenüber den Mitteln, mit deren Hilfe solche Verallgemeinerungen möglich werden. Sprachliche Probleme, die auf der Arithmetik gründen, jedoch einen logischen Symbolismus implizieren, und die von Anfang an ein Ordnen der Beziehungen verlangen, verursachen entscheidende Schwierigkeiten" (Bernstein 1967, S. 362). Man folgerte daraus eine Priorität sprachlicher Einflußnahme auf die Intelligenzentwicklung und sah sich indirekt durch die vermeintlich retardierte Intelligenzentwicklung Gehörloser bestätigt. Dagegen wäre einzuwenden, daß es sich bei diesbezüglichen Nachweisen unseres Wissens immer um sog. *allgemeine* Intelligenzbestimmungen handelte, deren Testunterlagen also auf der Spearman-schen Zweifaktoretheorie der Intelligenz (unter Einschluß verbaler Testaufgaben) basierten. Neuere – mehr oder weniger nonverbale – Intelligenzprüfungen erbrachten keineswegs durchgängig (in allen Intelligenzdimensionen sub specie multifaktorieller Modellvorstellungen; vgl. auch die Ergebnisse im empirischen Teil dieses Buches) die erwartete Leistungsminderung Hörsprachgeschädigter, wenngleich Besonderheiten des Intelligenzleistungsprofils Tauber und hochgradig Schwerhöriger unverkennbare Folgen sprachlicher Deprivation aufweisen. Begriffsbildungsprozesse und Klassifikationsleistungen, abstraktes, kausalanalytisches und schlußfolgerndes Denken, vor allem aber kreative Leistungen des divergierenden Denkens, des Problemlösungsverhaltens oder auch im numerischen Bereich sind für Hörsprachgeschädigte erfahrungsgemäß vielfach erschwert. Nach Jensen (1967) zeigt sich aber der negative Einfluß sozio-kulturell verursachter Sprachverzögerung im Hinblick auf die intellektuelle Entwicklung erst ab dem 4. Lebensjahr deutlicher, was hier mit dem Hinweis mangelnder verbaler Vermittlungskodes, die gerade für kognitive Planungs- und Abstraktionsprozesse als unerläßlich angesehen werden, erklärt wird. Gestützt wird diese Interpretation durch die Tatsache, daß spezifisch menschliche Intelligenzleistungen, wie begriffliches Lernen, abstraktes Denken (besonders im semantischen Bereich), Erarbeitung von Problemlösungsstrategien etc., in der Regel sich nicht vor dem 4. Lebensjahr manifestieren (vgl. Hofstätter 1954), d.h. erst ab dem Zeitpunkt offenbar werden, zu dem sich die menschliche Sprache lexikalisch sowie grammatisch-syntaktisch in der Ontogenese voll ausbilden kann. Vorher werden bei Intelligenzprüfungen (s.S. 32f.) überwiegend nur elementare (sprachfreie) Funktionen, z.B. Kurzzeitgedächtnis bzw. sog. Aufmerksamkeitsspannen sowie mehr assoziative Lernleistungen, erfaßt. Jensen verzeichnete nun in den unteren Sozialschichten mit zunehmendem Alter (etwa ab 5. Lebensjahr) eine Divergenz von elementaren und symbolisch vermittelten Intelligenzleistungen, während dieser Trend bei Mittelschichtkindern nicht festzustellen war. Weiterhin paßt in dieses Bild, daß korrelative Zusammenhänge zwischen Intelligenz und Sozialstatus erst *nach* dem 3./4. Lebensjahr nachzuweisen sind. Sämtliche aufgezählten empirischen Befunde verweisen auf das Gewicht sprachlicher (i.w.S. sozio-kultureller) Einflüsse auf die intellektuelle Entwicklung des Menschen.

### *Die Sapir-Whorf-Hypothese*

Nachdem also die Tatsache der Milieuwirkung und im besonderen hierbei die Rolle der Sprache als Einflußvariable der Intelligenz feststehen, erhebt sich jetzt die (mehr theoretische) Frage, *wie* sich im einzelnen wohl dieser Einfluß der Sprachstruktur auf die kognitiven Funktionen vollzieht. Eine der bekanntesten Hypothesen hierzu wurde von Sapir und Whorf aufgestellt und hat inzwischen z.T. unterschiedliche Auslegungen erfahren. Ihre grundlegende – freilich nicht uneingeschränkte – Bedeutung erweist sich u.E. gerade auch im Hinblick auf die Kognitionsvorgänge bei Gehörlosen, weshalb eine kurze Explikation in Anlehnung an die Oevermannschen Ausführungen (1969, S. 320 ff.) hier eingeschaltet wird.

„Von der Warte der Sprache aus gesehen, kann man das Denken als das stärkste Konzentrat betrachten, das die Sprache hergibt, wenn man jedes der Elemente eines normalen Sprechakts auf seinen Begriffsinhalt hin ausschöpft. Daraus ergibt sich ohne weiteres, daß Sprache und Denken nicht völlig identisch sein können. Außerstenfalls kann die Sprache als die nach außen gekehrte Seite des Denkens verstanden werden, und zwar auf dem höchsten abstrakten Niveau, wo die symbolischen Ausdrucksformen zu Hause sind. Meiner eigenen Überzeugung nach ist die Sprache im wesentlichen eine prärationelle Funktion. Sie arbeitet sich sozusagen ganz bescheiden in die Höhe bis zu dem Punkt, wo das Denken, das als latente Möglichkeit in den Kategorien und Formen der Sprache vorhanden ist, schließlich aus diesen Kategorien und Formen herausgelesen werden kann. Auf keinen Fall besteht, wie naiverweise oft angenommen wird, die Funktion der Sprache darin, bereits fertige Gedanken mit einem Namensschild zu versehen“ (Sapir 1961, S. 22. f.).

Mit den letzten Ausführungen wird die extreme Position sprachinstrumentalistischer Auffassungen klar zurückgewiesen. Dies kommt im folgenden Zitat noch deutlicher zum Ausdruck.

„Die Formulierung von Gedanken ist kein unabhängiger Vorgang, der im alten Sinne dieses Wortes rational ist, sondern er ist beeinflußt von der jeweiligen Grammatik... Wir gliedern die Natur in Linien auf, die uns durch unsere Muttersprache vorgegeben sind. Die Kategorien und Typen, die wir aus der phänomenalen Welt herausheben, finden wir nicht einfach in ihr – etwa weil sie jedem Beobachter in die Augen springen; ganz im Gegenteil präsentiert sich die Welt in einem kaleidoskopartigen Strom von Eindrücken, der durch unseren Geist organisiert werden muß – das aber heißt weitgehend: von dem linguistischen System in unserem Geist“ (Whorf 1963, S. 12).

Die Hypothese von Sapir und Whorf enthält zweifellos Aspekte erheblicher Evidenz. Konsequenterweise zu Ende gedacht würde sie allerdings implizieren, daß menschliche Erkenntnisse in ihrer Gänze oder doch wenigstens prinzipiell erst durch die Sprache ermöglicht würden, was wohl eine allzu kühne und vorschnelle Behauptung bedeutete. Auch ließe sich eine derart strenge Hypothesenauffassung kaum empirisch verifizieren bzw. falsifizieren, da ja dann (laut Hypothesenaussage) die kognitiven Prozesse nicht in sprachfreie versus sprachgebundene differenziert, d.h. unabhängig voneinander gemessen werden könnten. So ist wohl eine „weiche“ Interpretation der Hypothese angebracht. Oevermann referiert zwar eine Reihe experimenteller Untersuchungen, die das Sapir-Whorf-Thesegebäude eher bestätigen als widerlegen, doch lassen sich auch plausible Gegenargumente gegen die *sprachliche Determinismus-Hypothese* resp. die *Hypothese der linguistischen Relativität* ins Feld führen (siehe u.a. Hörmann 1967, bes. S. 334 ff., sowie Oevermann 1969, S. 328 ff.), ohne daß wir hierauf im einzelnen eingehen können<sup>29</sup>. Im Kern berührt diese Theorie – wenigstens in ihrer ersten These – unverkennbar Humboldtsche Gedankengänge.

Andererseits ist mit dieser Kritik noch nicht die Frage nach sprachunabhängigen Kognitionsleistungen irgendwie positiver beantwortet. Unter Bezug auf einschlägige empirische Forschungsansätze meint Oevermann hierzu: „Offensichtlich wirkten die linguistischen Elemente bei der äußerlich sprachfreien Aufgabenlösung spontan als innere Vermittlungsglieder und kanalisieren so die kognitiven Prozesse. Diese Interpretation finden wir in der *Hypothese der Sprachschleife* bei Glanzer/Clarac (1964): Zwischen Stimulus und Response tritt in der Wahrnehmung eine innere, stumme Verbalisierung, die die Reaktionen steuert. Glanzers Experiment

---

<sup>29</sup> Die *These des linguistischen Determinismus* besagt ja, daß die Sprache (schließlich) das Denken bestimme. Die *These der linguistischen Relativität* will zum Ausdruck bringen, daß jede Sprache eine bestimmte, d.h. hier relativ gültige, Welt-Anschauung bedinge. Sofern die erste These gilt, folgt notwendig daraus die zweite.

bestätigt indirekt diese Hypothese... Wir müssen annehmen, daß die Funktion der sprachlichen Symbolorganisation umso größer wird, je komplexere und schwierigere Probleme zu lösen sind. Erst dabei dürfte die Speicherung und Planung komplexer Stimulus-Response-Sequenzen mit Hilfe der Sprache sichtbar werden" (loc. cit., S. 322 f.).

Gegen einen direkten Kausalzusammenhang von Sprache und Kognition scheint vor allem eine Reihe von Begriffsbildungsexperimenten in ihren Ergebnissen zu sprechen. So können offenbar Kinder durchaus Kategorisierungsregeln logisch erfassen und anwenden, obwohl sie die betr. Begriffe falsch bzw. widersprüchlich verbalisieren, wie Piaget u.a. zeigen konnten; siehe dazu auch Hudelmayer 1970, S. 166 f. Nach Piaget ist die Sprache zwar eine notwendige, jedoch keine hinreichende Bedingung für den Aufbau logischer Operationen. Beide werden vielmehr in einem reziproken Verhältnis gesehen, wobei der Intelligenz die Rolle der unabhängigen Variablen zukommt. Siehe auch die älteren Untersuchungen von Ziehen (1933) und Hollstein (1953).

### *Die Kognition Hör-Sprachgeschädigter*

Die Analyse der *Kognition Gehörloser bzw. Hörsprachbehinderter* bietet sozusagen das klassische Untersuchungsfeld im Hinblick auf die zuletzt erörterte Problematik. Oevermann zitiert in diesem Zusammenhang bei uns wenig bekannte Untersuchungen anglo-amerikanischer Forscher (z.B. Oléron/Herren 1961, Goetzinger/Rousey 1959, Templin 1950, Wooden 1963, Lawton 1968 u.a.). Sie erbrachten übereinstimmend teilweise erhebliche Retardierungen im kausalanalytischen und abstrakten Denken. „Diesen Ergebnissen stehen neuerdings die Untersuchungen von Lenneberg (1967) und Furth (1966) entgegen, in denen in sprachfreien kognitiven Tests auch bei jenen gehörlosen Kindern keine wesentliche Retardierung festgestellt wurde, bei denen eine Sprachanbahnung kaum oder nur sehr oberflächlich stattgefunden hatte. Furth führt diesen Widerspruch darauf zurück, daß in den meisten früheren Untersuchungen nicht sorgfältig genug darauf geachtet worden sei, daß die experimentellen Bedingungen auch tatsächlich sprachunabhängig waren<sup>30</sup>. Dadurch sei die „Retardierung“ der Gehörlosen ex definitione eingeführt worden. Er selbst stellte in zahlreichen Vergleichen von gehörlosen und normalen Kindern in mehreren Altersstufen fest, daß keine Unterschiede in folgenden Aufgaben bestanden: Klassifikationen nach dem Prinzip der Gleichheit, Ähnlichkeit und Symmetrie; Auswahl von Bildern nach dem Beziehungsprinzip von Teilen und Ganzen; Behalten einer Folge von Formen; Organisation visueller Wahrnehmungen von Gestaltprinzipien; Anwendung abstrakter logischer Klassifikationsprinzipien. Von besonderem Interesse ist, daß die gehörlosen Kinder fast ebensooft „reversalshift“-Responses wählten wie normale Kinder. Schlechtere Leistungen zeigten die Gehörlosen im Erfassen von Konservierungskonzepten und allgemein in der Induktion und Entdeckung logischer Klassifikationsregeln. Furth erklärt jedoch diese Unterschiede nicht durch das Fehlen einer sprachlichen Symbolorganisation an sich, sondern durch die dadurch verursachte *geringere Umweltanregung*. Der Einfluß der Sprache auf die Kognition wird hier also implizit über ihre Funktionen in der sozialen Kommunikation gedeutet.

Nun läßt sich zwar gegen Furth einwenden, daß der Zwang zur sprachfreien experimentellen Anordnung die kognitiven Aufgaben künstlich vereinfachte, daß gerade die Retardierung in der Fähigkeit, logische Klassifikationsregeln zu entdecken, und die allgemeine Stimulusrigidität die Reduzierung von Lernmöglichkeiten zeigen, daß bisher viele Untersuchungen eine Retardierung, aber keine einzige eine kognitive Überlegenheit gegenüber normalen Kindern belegten (Furth 1964), daß schließlich gehörlose Kinder zumindest in einer durch Sprache organisierten

<sup>30</sup> Auf die Problematik nonverbaler Intelligenztestung kommen wir später zurück (vgl. Neumann 1968).



sozialen Umwelt aufwachsen und daß sie, da ihnen die Sprache von vornherein verschlossen bleibt, andere, der sprachlichen analoge Symbolorganisationen ausbilden. Aber alle diese Argumente können die These einer intrapsychischen direkten Kausalbeziehung zwischen der Verfügung über das sprachliche Zeichensystem und den kognitiven Funktionen angesichts der einheitlichen Tendenz in Furths Ergebnissen nicht retten. Würde sie zutreffen, so dürften nicht in so vielen Fällen mit linguistischer Kompetenz ausgestattete Kinder schwächere kognitive Leistungen als gehörlose Kinder zeigen" (Oevermann 1969, S. 327 f.).

Mit diesen Feststellungen soll die Bedeutung der Sprache im Hinblick auf die Organisation vorab komplexer kognitiver Prozesse keinesfalls in Frage gestellt werden. Im Sinne der Whorf-Hypothese ist hierbei der *syntaktischen* Organisation vor der lexikalischen Gliederung das Hauptgewicht zuzuerkennen. Sehr wahrscheinlich ist weniger das sprachliche Symbolsystem als solches als vielmehr dessen *aktiver Einsatz* in kognitiven Prozessen von entscheidendem Einfluß, gerade auch im Hinblick auf die Entwicklung von Begabung und Intelligenzleistungen. Ein direkter kausaler Zusammenhang von Sprache und Kognition gilt nach den neuesten Forschungsergebnissen als unwahrscheinlich. Andererseits spielen auch nichtsprachliche „Vermittler“ eine Rolle in Kognitionsvorgängen (Bruner et al. 1966). „Aber die Sprache zeichnet sich vor allen anderen Symbolsystemen durch ihre konsistente, hierarchisch gegliederte Regelmäßigkeit aus. Das macht sie zu einem höchst wirksamen Vehikel der Begriffsbildung" (Oevermann 1969, S. 325) resp. – so können wir weiter generalisieren – im Aufbau intelligenter Verhaltens.

In unserer bisherigen Erörterung sozio-kultureller Determinanten von Begabung oder Intelligenz <sup>31</sup> überwog das außer- bzw. vorschulische Erziehungs- und Sprachmilieu bei weitem. Das erklärt sich vor allem aus der inzwischen recht gut erhärteten Annahme, daß gerade die ersten 6 Lebensjahre und damit das familiäre und außerschulische Erziehungs- und Sozialfeld für den Aufbau der Intelligenz(-Leistung) wie keine Epoche nachher entscheidend sind. Daß der Schule und hier besonders den Lehrern darüberhinaus durch sozialpädagogische Aufgeschlossenheit, durch methodisch-didaktische Unterrichtsmaßnahmen, durch Curricula-Auswahl u.ä. noch erhebliche Spielräume des Begabens in die Hand gegeben sind, hat wohl nie jemand ernsthaft bestritten. Im Hinblick auf die sozialkulturelle, vorab sprachliche Deprivation vieler Unterschichtkinder fällt hier der Schule in erster Linie die Aufgabe *kompensatorischer* Bildungsbemühungen zu. „Der Umstand, daß das Kind aus der Arbeiterklasse einem anderen Aspekt der Sprache Bedeutung beimißt als dem, der vom Schulunterricht verlangt wird, ist für den Widerstand verantwortlich, den es der Erweiterung des Vokabulars, der Handhabung von Wörtern und der Konstruktion von geordneten Sätzen entgegensetzt, *was schließlich zu einer allgemeinen kognitiven Verarmung führt*" (Bernstein 1967, S. 362). Hier kann der Lehrer, der selbst die Sprache der Mittel- und Oberschicht gebraucht und quasie eine Mittelstands-Ideologie – zumindest in Bildungsfragen – vertritt, dem Arbeiter- und Bauernkind durch sein Wissen um die verschiedenen Sprachkodes und Erziehungsstile sowie sein Verständnis für die Schwierigkeiten einer sprachlichen Adaptation (des restringierten an den elaborierten Kode) resp. vermeintlich fehlende Bildungsinteressen und Lernmotivation sehr entscheidend zur Begabungsförderung beitragen. Wie dies im einzelnen geschehen kann, wird Gegenstand der nachfolgenden Erörterungen sein. Nach einer Studie von Newman et al. (1937), die sich auf 19 in unterschiedlichem Milieu aufgewachsene EZ-Paare bezieht, korrelierten sozial-

---

<sup>31</sup> Daß auch nichtkognitive *Persönlichkeitsbedingungen* (z.B. Ängstlichkeit, Rigidität, Neurotizismus) die Intelligenzleistung beeinflussen, wurde schon früher erwähnt. Um unsere theoretischen Erörterungen nicht zu sehr auszuweiten, verzichten wir hier auf eine Erklärung dieser Persönlichkeitsvariablen, verweisen aber den interessierten Leser auf einen kürzlich erschienenen Aufsatz von uns (Heller 1969a); siehe auch Heller (1970a) und Lewis (1963). Motivation und Interessen sind natürlich ebenfalls als Persönlichkeitskomponenten anzusprechen.

pädagogische Kriterien am höchsten mit den Schulleistungsfähigkeiten (+ 0.91), während die zweithöchste Korrelation (+ 0.79) zum Binet-IQ ermittelt wurde.

Diese Befunde verweisen noch auf eine andere Problematik, die darin zu liegen scheint, daß der in der Regel aus der Mittelschicht kommende Lehrer allzuleicht geneigt ist, seine „Begabungsmaßstäbe“ bei der *Beurteilung* der Schulleistungen anzulegen. „Nicht nur schulische Leistung, sondern die Forderung des Lehrers nach besonderen Qualitäten, die seinem Erziehungsideal entsprechen, werden als Maßstab für eine Begabung und einen Bildungsweg angesehen“ (Gaus 1969, S. 178). Solche vielleicht etwas überspitzt ausgedrückte Kritik am vielfach subjektiven schulischen Bewertungssystem mag manchen Leser befremden. Immerhin kommt Latscha aufgrund eigener empirischer Untersuchungen (schweizer Verhältnisse) zu folgender Aussage: „Die Noten und Empfehlungen des Primarlehrers werden in hohem Maße von typologischen Eignungsurteilen mitbestimmt, in die sehr oft unreflektiert milieubedingte Verhaltensmerkmale eingehen“ (1966, S. 257).

„Die Schichtzugehörigkeit einer Familie bedeutet für die ‚*Begabungsentwicklung*‘ des Kindes, daß durch die Sozialisation in der Familie Werte und Verhaltensweisen vermittelt werden, die je nach Schicht in unterschiedlicher Beziehung zu den von der Gesellschaft gesetzten Begriffsinhalten von ‚Begabung‘, Intelligenz und Schulerfolg stehen... Die Schule als ‚Mittelklasseinstitution‘... begünstigt mit ihren Forderungen nach adäquaten Verhaltensweisen und nach einer bestimmten Form der Leistung (theoretisch, geistig) diejenigen Schüler, die die vorausgesetzten Werte bereits in der Familie erlernen können und dauernd gestützt werden durch ein entsprechend strukturiertes familiäres Milieu... Die höheren Schulen vertreten (z.B.) einen Bildungsbegriff, der traditionell der Oberschicht zugehörte und an dem die niedrigeren Schichten keinen Anteil hatten. Als ‚echte‘ Bildung gilt auch heute noch ziemlich verbreitet nur jene literarisch-ästhetisch-philologische, und diese gehört zu den spezifischen Normen der Mittel- und Oberschicht.

Es ist einsichtig, daß den unteren Schichten solch ein Begriff fremd ist; sie werden diese spezifischen Inhalte als für sich selbst nicht möglich und erreichbar ansehen, oder sie werden diese Bildung als ihrem sozialen Bewußtsein inadäquat ablehnen. Von ihrer Lebenserfahrung und Soziallage her richtet sich ihr Interesse mehr auf technische, ökonomische, politische und soziale Fragen als auf literarisch-geistige Inhalte“ (Gaus, loc. cit. S. 181 f.)

Ohne hier die Problematik der heute viel diskutierten „Ungleichheit der Bildungschancen“ in ihrem sozialpolitischen Kontext weiterzuverfolgen, dürfte doch die begabungsfördernde (versus hemmende) Bedeutung des schulischen Erziehungs- und Bildungsmilieus in wesentlichen Ausschnitten transparent geworden sein. Natürlich stehen nach wie vor Denkleistungen bzw. Lernfähigkeiten als wichtigste kognitive Bedingungen im Zentrum schulischer Bildungsbemühung. Die Kenntnis darüber, daß diese Funktionen weithin – gerade auch vom Lehrerverhalten – abhängige Variablen darstellen (vgl. noch Heller 1969a, 1970a), sollte den Pädagogen zu einer optimalen Begabungsförderung (besonders sozial und sprachlich benachteiligter Kinder) ermutigen. Im nächsten Abschnitt seien deshalb bildungspolitische Vorschläge sowie mehr praktisch-pädagogische Maßnahmen zur Aktivierung von Begabung und Lernleistung diskutiert.

### 3. Pädagogische und bildungsrelevante Folgerungen

*Kompensatorische Erziehungsmaßnahmen* und *Modifikationen schulischer Leistungsbeurteilung* stellen gleichermaßen mittelbare Ziele zur Beseitigung sozio-kulturell bedingter Begabungsdifferenzen dar. Folgende konkrete Schritte werden hierzu nötig sein.

Bereits im *Vorschulalter* kommt es darauf an, die durch Erziehungsstile, Sprachverhalten u.ä. restringierten Erfahrungsmöglichkeiten des Unterschicht-Kindes so zu erweitern, daß der spätere Schuleintritt nicht einem „Milieubruch“ gleichkommt. „Die Schule sollte nicht versuchen, den *öffentlichen* Sprachgebrauch auszurotten, der ja nicht nur seine eigene Ästhetik besitzt, sondern den Sprechenden auch psychisch mit seinesgleichen und seinen lokalen Traditionen verknüpft. Es handelt sich vielmehr darum, diesen Sprachgebrauch

zu ergänzen. Die schulische Entwicklung des Kindes aus der Arbeiterklasse hängt in erster Linie von Bedingungen und Inhalt der Erziehung in Kindergarten und Elementarschule ab" (Bernstein 1967, S. 364). Dies bedeutet, daß ausgleichende Bildungsbemühungen von seiten des Vor- und Grundschullehrers, der Kindergartenpädagogen usw. zunächst auf Erziehungsgewohnheiten und typische Sprachstile der Unterschichtkinder eingehen, diese als (vorläufige) Erziehungs- und Kommunikationsmittel annehmen und erst allmählich spezifische Modifikationen ins Auge fassen sollten. Die Ermöglichung entsprechender kognitiver Erfahrungsfelder, deren Erweiterung zunehmend gezielter und systematischer zu betreiben wäre, kann – unter Vermeidung negativer Sanktionen – so dazu führen, daß von den Kindern selbst her, etwa in bestimmten Problemsituationen, die Einsicht in die größere Leistungsfähigkeit formal und inhaltlich elaborierter Sprachmittel wächst oder doch zumindest durch probierendes Handeln an konkreten Lösungsobjekten Vorteile des formalen Sprachkodes (mehr unbewußt) erfahren werden.

In der Anfangsphase wird beispielsweise der Pädagoge darauf achten müssen, die verhältnismäßig „undifferenzierte, distanzlose Ausdrucksweise im ‚restringierten Kode‘ nicht nach Kriterien mittelständischen Sprachgebrauchs als von den Kindern intendierte Verletzung mittelständischer Regeln des Wohlverhaltens (zu) deuten. Die Tendenz der Unterschichtkinder, wenig und nur knapp zu verbalisieren, darf bei der Leistungsbeurteilung nicht unbedingt als mangelndes Interesse am Lehrstoff interpretiert werden. Sie entspricht der Planungsstrategie im ‚restringierten Kode‘. Schließlich dürfen die Inhalte der Spracherziehung bei Aufsätzen und Lesestücken nicht wie bisher weitgehend außerhalb des Erfahrungsmilieus dieser Kinder liegen. Die seit längerem geforderte Lesebuchreform erhält unter diesem Gesichtspunkt eine zusätzliche Bedeutung" (Oevermann 1969, S. 340).

Die Verwirklichung dieser Ziele wird – schon unter Hinweis auf mögliche Überforderung gerade seitens von Unterschicht-Eltern – sehr oft mit dem Postulat *vorschulischer Bildungseinrichtungen* verknüpft. Eine solche „Vorschule als Anregungsinstitution" könnte nach Heckhausen (1969, S. 207 f.) eine frühkindliche Entwicklungsförderung sicherstellen, „die das durchschnittliche Maß des in den Elternhäusern Erreichten übersteigt sowie insbesondere das Defizit an Anregungspotential in den unteren Sozialschichten und jenen Sozialgruppen ausgleicht, deren sozio-kulturelles Milieu die Entfaltung der kindlichen Potentialitäten schon in den ersten Lebensjahren vermindert. Das betrifft vor allem die großstädtischen Wohngegenden der unteren Sozialschichten sowie auch die ländlichen Streusiedlungen und Kleingemeinden. Die Vorschule soll nicht die Familie ersetzen, sondern nur in einigen Bereichen ergänzen. Sie soll auch anregend in die Familie hineinwirken. Deshalb ist eine weite Öffnung zur Familie hin, ja, die Einbeziehung von Müttern als Beobachter und gar Mitwirkende in Vorschulprogrammen wünschenswert. Mütter können dadurch fördernde Ansteckungswirkungen in ihre Familie, ja sogar in die Nachbarschaft hineintragen (Gray/Klaus 1966)... Der ‚Lehrplan‘ sollte in erster Linie auf zwei Ziele ausgerichtet sein: auf eine Beschleunigung der sachstrukturellen Entwicklung und auf die Funktionsübung zweckdienlicher kognitiver Stile. Geschieht das in entwicklungsgerechter Weise, so würde damit gleichzeitig die Ausbildung der Leistungsmotivation und der Wertschätzung von tüchtigkeitsbezogenen Sachbereichen gefördert."

Unter *sachstrukturelle Entwicklungsprogramme*<sup>32</sup> fallen vor allem Maßnahmen zur *Weckung des Symbolverständnisses* und zur *Sprachförderung*: Zuordnungsübungen von Schriftsymbolen und Sachgegenständen (etwa im 4./5. Lebensjahr) bzw. Vorbereitung des Lesenlernens nach einschlägigen Prinzipien (siehe u.a. die Modelle von E. Kern 1958, Schulte 1968, A. u. E. Kern 1969, Doman/Lückert 1967, Kratzmeier 1970; vgl. auch

<sup>32</sup> „Unter *sachstrukturellem Entwicklungsstand* sind die Kenntnisse und Fertigkeiten eines Schülers zu verstehen, die er zu einem gegebenen Zeitpunkt seiner Entwicklung im Hinblick auf den relevanten Sachbereich der gegebenen Unterrichtssituation besitzt. Dazu gehören insbesondere Wissenselemente, Sprachumfang, Verknüpfungsregeln, sensumotorische Fertigkeiten" (Heckhausen 1969, S. 193, vgl. auch S. 199 ff.). Zur weiteren Bedeutungsexplikation siehe noch Aebli 1969.

Schmalohr 1971, Getzels 1966 und Hess 1965), Übungen zur Artikulations- und Lautdiskriminierfähigkeit (phonematische Analyse) sowie quantitative und qualitative Wortschatzerweiterung bis in die Grundschulzeit hinein (vgl. Schulte 1969), Anschauungsübungen (vgl. J. Wittmann 1933), Förderung sprachlicher Differenzierung nach der grammatisch-syntaktischen Seite hin u. dgl. m.

Neben der „sachimmanenten Entfaltung logischer Strukturen im Begreifen und Wissen“ bedarf es nach Heckhausen (loc. cit.) der *Einübung von Operationsmodi*, soweit sie kognitiven und intellektuellen Prozessen zugrunde liegen; „ja, es kommt wahrscheinlich sogar mehr auf die verfügbaren Operationsmodi an als auf die erworbenen intellektuellen Fertigungsprodukte“. Hierzu könnten entwicklungsgemäß zusammengestellte spielerische Funktionsübungen auf der Basis diverser kognitiver Operationsmodi beitragen: Aufmerksamkeitstraining, Wahrnehmungsübungen verschiedenster Art (Hören, Sehen, Tasten) sowohl „passiv-global“ (feldabhängig) als auch „aktiv-analytisch“ (feldunabhängig), Aufgaben zu impulsiver versus reflexiver Problemlösung, Erarbeitung von Lernstrategien („Lernen zu lernen“) u.ä.

In den genannten intellektuellen Fördermaßnahmen sowie dem Einspielen kognitiver Operationen sieht Heckhausen zugleich wichtige Modi der „Verstärkung einer erfolgszuversichtlichen Leistungsmotivation“. Dabei sollte die *intrinsische* Lernmotivierung, d.h. sachbereichsbezogene Anregungsvariablen, gegenüber den Möglichkeiten *extrinsischer* Motivierung (z.B. durch Aktivierung persönlicher Identifikationsbedürfnisse, etwa mit dem Lehrervorbild, oder durch Strafandrohung) bevorzugt werden.

„Die wichtigste Bedingung für eine intrinsische Lernmotivierung des Kindes und damit für das Gelingen frühkindlicher Förderungsprogramme ist ... die Schaffung motivierender Diskrepanzerlebnisse. Die gestellte Aufgabe muß, um Aufmerksamkeit und tätige Bemühung des Kindes auf sich zu ziehen, einen dosierten Schwierigkeitsgrad haben, der den sachstrukturellen Entwicklungsstand des Kindes um ein geringes übersteigt. Momentaner Fähigkeitsstand und Aufgabenanforderung müssen fortlaufend aufeinander passen, zwischen beidem muß ‚Passung‘ sein (match, vgl. Hunt 1964). Das ist am besten gewährleistet, wenn das Aufgabenmaterial eine Schwierigkeitsgraduierung zuläßt, die das Kind selbst manipulieren kann; oder wenn das Kind Material mit dem jeweils passenden Schwierigkeitsgrad selbst auswählen kann. Dies setzt eine individualisierte Selbstbeschäftigung jedes einzelnen Kindes voraus. Ihr ist in der Vorschule deshalb über weite Strecken der Vorzug gegenüber einer Gruppenbeschäftigung einzuräumen, weil sonst die optimale Passung, die das einzelne Kind am besten für sich selbst finden kann, verlorengehen muß.

Die Montessori-Pädagogik hat dieses Prinzip der intrinsischen Motivierung durch selbstgesteuerte Eigentätigkeit des Kindes in hohem Grade kultiviert. Manches könnte von ihr übernommen werden. Die landläufige Kindergartenpädagogik scheint dagegen wenig geeignet, die Möglichkeiten einer optimalen Wechselwirkung von intrinsischer Motivation und Funktionsübung auszuschöpfen. Der Kindergarten ist immer noch zu sehr ‚Bewahranstalt‘. Im Geiste Fröbels gibt es viel zu tun, um die neueren, entwicklungspsychologischen Erkenntnisse in die Anwendungspraxis einer neuen Vorschuldidaktik umzuwandeln“ (Heckhausen 1969, S. 209).

Daß solche vorschulischen Erziehungs- und Bildungsprogramme vor allem im Hinblick auf die Bedürfnisse sozio-kulturell benachteiligter Kinder erstellt werden sollten, haben wir schon verschiedentlich, zuletzt im Anschluß an die Erörterung der Sapir-Whorf-Hypothese, betont, Hauptziel vorschulischer Förderung, in das letztlich alle Bildungsbemühungen einmünden, wäre demnach die Vermittlung syntaktischer (logischer) Strukturen oder Regeln qua kognitiver Stimulations-Modi der Weltzuwendung. Dies ist in dieser frühen Phase vorab durch Konfrontation mit einem differenzierten, d.h. lexikalisch und besonders grammatisch-syntaktisch elaborierten, Sprachmodell möglich. Solche Strukturmodelle Kindern aus sprachlich restringierten Familienverhältnissen zu bieten, wird deshalb zu einer der anspruchsvollsten Aufgaben vorschulischer Betreuung. Nach Oevermann genügt jedoch ab dem 5./6. Lebensjahr eine „bloße Konfrontation mit einem differenzierten Sprachmodell“ nicht mehr. „Zwei Elemente müssen hinzutreten: a) Die Kinder müssen angehalten werden, ihre sozialen Interaktionen und individuellen Problemlösungs-

versuche beispielsweise in klassifikatorischen Spielen und Aufgaben kontrolliert zu verbalisieren. b) Sie müssen vor allem in wechselnde, differenzierte Interaktionssituationen hineingezogen werden, um differenzierte Strategien der verbalen Planung im Handlungs-vollzug zu lernen" (1969, S. 341). Beispielhaft werden kooperative Problemlösungen Koalitionsbildungen, „Verhandlungen“ zur Abstimmung divergierender Handlungsziele, Dialoge mit dem „Lehrer“ bei wechselnden Rollenübernahmen u.ä. aufgezählt, kurz: es „müssen Interaktionskontexte geschaffen werden, die eine situative differenzierte verbale Planung provozieren" (loc. cit.).

Vom Standpunkt kognitiver bzw. intellektueller Optimalförderung sind sämtliche aufgezeigten Ansätze zweifellos sehr wichtig, selbst wenn man den vielfach durch vorschulische „Förderungseingriffe“ erzielten (und nicht weniger oft sehr naiv interpretierten) IQ-Gewinnen nicht die entscheidende Kriteriumsfunktion beizumessen gewillt ist. Andererseits muß aber auch auf nicht zu übersehende Probleme hingewiesen werden. Die Gefahr der Entfremdung des Kindes vom Verhaltensstil seiner Familie durch zunehmende (eigene) kognitive und sprachliche Strukturierung sowie die eng damit zusammenhängende Lockerung emotionaler und sozialer Bindungen stellt nur ein, aber vielleicht das gravierendste, Konfliktmoment dar. Die theoretische Forderung, die Eltern und hier besonders die Mütter aktiv zu beteiligen resp. in die Förderungsvorhaben zu integrieren, ist sicher richtig, praktisch dürfte dies allerdings noch wenig befriedigend oder eben nur bei Mittelschicht-Eltern gelungen sein. Und selbst da gewinnt man den Eindruck, als ob das Anliegen optimaler Begabungsförderung sehr oft mit einer Vorverlegung der Pflichtschulzeit bzw. Verlängerung der Leistungsschule nach „unten“ verwechselt würde. Eine weitere Problematik liegt in der Frage, wie man *konkret* eine solche Spracherziehung durchführen kann, d.h. in der Erarbeitung noch weithin desiderabler Modi der *Verzahnung des formalen (elaborierten) mit dem öffentlichen (restringierten) Sprachkode*. Darin möchten wir sogar das vorläufig schwerste Hindernis auf dem Wege zur Verwirklichung der in diesem Abschnitt aufgewiesenen Ziele vermuten. Den Forschungsinitiativen der nächsten Jahre öffnet sich hier eines der verdienstvollsten und zugleich dornenreichsten Projekte, das wahrscheinlich nur im „Action research“-Ansatz zu bewältigen sein wird.

Die Anstrengungen um eine (neue) *Vorschuldidaktik* sind deshalb so wichtig, weil zwischen dem 3. und 6. Lebensjahr bereits „a) entscheidende Prozesse im primären Spracherwerb stattfinden, b) die wichtige Phase der Integration des sprachlichen Zeichensystems mit der operativen Intelligenz beginnt und c) die für die kognitive und motivationale Entwicklung bedeutsame Rollenkonstellation im Elternhaus in dieser Phase für die Unterschichtkinder besonders prekär ist" (Oevermann, loc. cit.).

Sollen die erwarteten und möglicherweise erzielten vorschulischen Lernerfolge gerade der Unterschichtkinder nicht wieder verloren gehen, so ist ein nahtloser Übergang zur Grundschule unerlässlich. Theoretisch wie praktisch impliziert dies *didaktische Konsequenzen für den Grundschulunterricht* (vgl. Kemmler 1967, Flechsig et al. 1969). Sub specie kompensatorischer Bildungsbemühungen werden hier neben der quantitativen und qualitativen Erweiterung des Begriffswortschatzes und der Festigung resp. weiteren Elaboration grammatisch-syntaktischer Sprachstrukturen vor allem die Nutzung der in der Sprache gelegenen Möglichkeiten zur Gewinnung von Denkstrategien (etwa bei Problemlösungen) sowie organisatorische Maßnahmen (z.B. Sonderkurse im Schreiben und Lesen) stehen. Lese- und Schreibfunktion bekommen ja nun ihre eigentliche Bedeutung.

„Daneben wird jetzt aber die Fähigkeit des Lesens und Schreibens immer wichtiger, weil dadurch unabhängig vom unmittelbaren Handlungskontext in abstrakten, rein linguistischen Kontexten neue Bedeutungen entwickelt werden, die über die unmittelbare Erfahrung weit hinausgehen. ‚Schreiben‘ soll hier nicht so sehr die Kunst des Rechtschreibens, sondern die des ‚Übersetzens‘ der eigenen mündlichen Sprache auf eine neue Ebene der Symbolorganisation bedeuten. Sonderkurse sind hier für die Kinder der Unterschicht möglicherweise nötig, wenn auch durch die Einrichtung der vorgenannten Maßnahmen schon ein Teil der Schwierigkeiten ausgeräumt werden kann. Die skizzierten Förderungsmaßnahmen müssen in allen Fächern, nicht nur in den sprachlichen, durchgehend Anwendung finden. Unab-

hängig von der traditionellen Fächereinteilung erscheint die regelmäßige „Unterweisung“ im kommunikativen und sprachlich-individuellen Problemlösen während der ersten Grundschuljahre empfehlenswert.“

Schließlich plädiert Oevermann für eine Revision des traditionellen Grammatikunterrichts in der Schule.

„Ein erfolgreicher Grammatikunterricht ist jeweils schon an die Beherrschung der syntaktischen Regeln, die jetzt auf einer zweiten Ebene explizit gemacht werden sollen, gebunden. Daher ist diese praktische Beherrschung vorrangiges Unterrichtsziel. Sie wird jedoch durch Unterweisung in der traditionellen Schulgrammatik mit ihren Ausnahmen und Besonderheiten kaum gefördert, ja, kann dadurch sogar gehemmt werden. Es müßte daher ein an der modernen generativen Grammatik orientierter Unterricht entwickelt werden, in dem die Sicherheit in der Verfügung über syntaktische Regeln gestützt *und* gleichzeitig das Regelgebäude der Sprache als Gegenstand der Erkenntnis rational und nicht nur normativ transparent gemacht wird. Eine Transformationsgrammatik Chomskyscher Prägung muß daher dringend für die deutsche Sprache geschrieben und für Unterrichtszwecke adaptiert werden“ (loc. cit., S. 341 f.).

Nach diesen sehr dezierten Hinweisen (siehe auch die eingehendere Darstellung bei Hörmann 1967, bes. S. 49 ff. und 248 ff.) an die Sprachwissenschaftler, die Sprachpsychologen und die (allgemeinen) Sprachpädagogen wäre nunmehr unter sonderpädagogischen Gesichtspunkten die Frage nach einschlägigen Parallelen zur Situation Hörsprachbehinderter zu stellen. Daß hier zahlreiche Berührungspunkte oder gar Entsprechungen vorliegen, dürfte unschwer zu erkennen sein. Um Wiederholungen möglichst zu vermeiden, seien am Ende des 1. Kapitels nur einige besonders aktuell erscheinende Problemaspekte angesprochen, zumal sich im empirischen Teil noch genügend Ansatzpunkte sonderpädagogischer Relevanz ergeben werden.

#### 4. Sonderpädagogische Perspektiven

Hinsichtlich des Sprachverhaltens unterscheiden sich *hörgeschädigte Kinder* im Intergruppenvergleich recht deutlich von den normalsinnigen Alterskameraden – auch bei kontrollierten Intelligenzverhältnissen. Andererseits sind aber mehr oder weniger ausgeprägte Intragruppendifferenzen bezüglich laut- und schriftsprachlicher Leistungseffizienz feststellbar, sowohl unter wahrnehmungsdifferentiellen Gesichtspunkten (z.B. zwischen Gehörlosen, Resthörigen, Schwerhörigen) als auch aus anderen kognitiven und nichtkognitiven Gründen (z.B. sub specie sprachlicher Fähigkeit resp. allgemeiner Begabung, familialem Erziehungs- und Kulturmilieu, Internat versus Externat etc.). Daraus ergibt sich die zwingende Notwendigkeit, jedem Bildungsprogramm eine möglichst umfassende *diagnostische Untersuchung des Einzelfalles* voranzustellen. Nur so wird es möglich sein, individuell abgestimmte Förderungsmaßnahmen zu erarbeiten und damit für jedes einzelne Kind optimale Bildungsvoraussetzungen zu schaffen. Alle bisherigen schulischen Differenzierungsversuche zeigten u.E. recht eindrucksvoll, daß eindimensionale Kategorisierungskonzepte (etwa Differenzierung ausschließlich nach Hörvermögen versus Intelligenz) das Problem individuell angemessener Bildung bzw. sprachlicher Entwicklung Hörgeschädigter nicht befriedigend lösen können (vgl. Heller 1967 a).

Ein *Diagnostikum* für die spezifischen Belange Hörsprachgeschädigter – um einmal diese Gruppe hier zu favorisieren – im Hinblick auf die thematisierten Intelligenz- und Sprachleistungsdimensionen müßte nach den bisherigen Erörterungen interdependenter Zusammenhänge von Kognitionsvorgängen und Sprachverhalten die wichtigsten, auf Seite 9 bis 30 explizierten Begabungsvariablen erfassen: nonverbale resp. Handlungs- und sprachliche Intelligenzfaktoren der sog. Kernintelligenz wie auch der Stützfunktionen der Intelligenz (Leistungsmotivation, Interessen, Konzentration, Arbeitsorgfalt etc.). Dazu werden wir im empirischen Teil neuere Untersuchungsergebnisse referieren können. Die im Anhang beigegebenen Normen für Gehörlose und Schwerhörige werden die oben erwähnten dia-

gnostischen Aufgaben in wesentlichen Punkten erleichtern. Sie ermöglichen nicht nur eine Diagnose der kognitiven Ausgangslage („Startbedingungen“) jedes Schülers vor entscheidenden Bildungsmaßnahmen, mit Hilfe der aufgeführten Tests bzw. unter Benutzung der gruppenspezifischen Normdaten werden auch indirekt Bildungsfortschritte (versus Stationationen) objektiv, d.h. in Relation zur repräsentativen Schulklassen- oder Alterspopulation feststellbar. Solche Testkontrollen informieren mit anderen Worten auch über die Begabungseffizienz von Unterricht und didaktischen Maßnahmen des Lehrers. Ferner sind sie im Hinblick auf die Erfassung und Beurteilung der Lernfortschritte des Schülers nützlich, wenngleich dies nicht unmittelbares Anliegen von Intelligenz- und Begabungsuntersuchungen sein kann. Dies wäre vielmehr Hauptaufgabe einer standardisierten Schulleistungsmessung, wozu jedoch im Bereich der Hörsprachgeschädigtendiagnostik bislang die notwendigen Methoden fehlen. Man muß sich aber immer vor Augen halten, daß Begabung und Leistung keine völlig voneinander unabhängig meßbaren Größen darstellen, genau so wenig wie ihre Abhängigkeit von sozio-kulturellen Variablen übersehen werden darf.

Speziell im Hinblick auf die pädagogische Kontrolle *sprachlicher* Entwicklung und Förderung Hörsprachgeschädigter sind – neben den gängigen Intelligenz- und Begabungstests – folgende Diagnostika von Interesse:

1. Verfahren zur *Erfassung grundlegender psycholinguistischer Fähigkeiten*, d.h. individueller kognitiver Voraussetzungen zum Aufbau versus zur Entschlüsselung linguistischer Codes. Quasi auf dem untersten Level wären hier vor allem *Signifikationsprozesse* sowie *sensumotorische* Bedingungskomponenten resp. Fähigkeiten der *Artikulation und Lautdiskriminierung* zu testen. Standardisierte Untersuchungsverfahren dieser Art liegen im deutschsprachigen Raum nicht vor. Im angloamerikanischen Bereich hat sich seit 1961 „The Illinois Test of Psycholinguistic Abilities“ (ITPA) von Kirk/McCarthy, der 1968 in 2. verbesserter Auflage erschienen ist, bewährt. Der ITPA erfaßt in 12 Subtests psycholinguistische Fähigkeiten auf 3 Ebenen: Auditory Reception, Visual Reception, Auditory Association, Visual Association, Verbal Expression, Manual Expression, Grammatic Closure, Visual Closure, Auditory Sequential Memory, Visual Sequential Memory, Auditory Closure and Sound Blending. Diese werden den Dimensionen „Channels of Communication“ (auditory-vocal resp. visual motor), „Psycholinguistic Process“ (receptive, organizing, expressive) und „Levels of Organization“ (automatic bzw. representational) zugeordnet. Die theoretische Verwandtschaft zum bekannten Osgood-Modell sprachlicher Kommunikationssysteme<sup>33</sup> ist unschwer zu erkennen. Dem ITPA sind Altersnormen für das 2. bis 10. Lebensjahr sowie eine Umrechnungstabelle zur Schätzung des Intelligenz-Alters (sensu Stanford-Binet) beigegeben. Die Testautoren heben die Bedeutung *intraindividueller* Fähigkeitsdifferenzierung gegenüber interindividuellen resp. Gruppenvergleichen hervor, wenngleich auch letzterer Gesichtspunkt im Falle eines möglichen Testeinsatzes bei Hörsprachgeschädigten (nach entsprechender Adaptation für deutschsprachige Verhältnisse bzw. einer wünschenswerten und erforderlichen Modifikation im Hinblick auf die Hörgeschädigtendiagnostik) differentialpsychologisch von Interesse wäre.

2. *Wortschatztests* zur Ermittlung des Begriffsumfangs. Im empirischen Teil dieses Buches wird über Untersuchungen mit dem Frankfurter Wortschatztest (WST) bei Gehörlosen und Schwerhörigen zu berichten sein. Dieser Versuch stellt einen vorläufigen Ansatz zur objektiven Bestimmung des Wortschatzes hörgeschädigter Schüler dar, wobei wir uns der Problematik dieses Unterfangens (mit Hilfe eines für Normalhörende konzipierten Tests) durchaus bewußt sind. Der Untersuchung kommt auch mehr mittelbare

<sup>33</sup>Nach Osgood (1957) eignet sich Sprache im sozialen Kontext, der hier durch die Pole „Encoding“ (Sender/Sprecher) und „Decoding“ (Empfänger/Hörer) sowie das Bindeglied Mitteilung oder Nachricht gekennzeichnet ist. In der Modifikation des Anwendungsmodells (ITPA) repräsentieren „Expressive Process“ (Expression), „Receptive Process“ (Reception) sowie „Organizing Process“ (Association) oben aufgeführte Dimensionen.

Bedeutung zu im Hinblick auf die Neukonstruktion eines spezifisch für Gehörlose geeigneten Inventars zur Erfassung des (absoluten) Begriffs-Wortschatzes. Darüberhinaus sollten auf *lexikalischer* Ebene auch qualitative Aspekte, etwa mit Hilfe des *Diversifikationsquotienten*, berücksichtigt werden (s.S. 40). Neben der Bestimmung des relativen Gebrauchs der *Wortarten* wären *sprachliche Analogie- und Abstraktionstests*, wie sie teilweise in verbalen Intelligenzmeßreihen für Normalhörende schon vorliegen, unter Berücksichtigung der Besonderheiten Hörsprachgeschädigter ins Auge zu fassen.

3. Schließlich müßten auf *syntaktischer* Ebene Analysen sprachlicher Komplexität durchgeführt werden. Satzlängen, Gebrauch von Relativsätzen wie überhaupt die Anzahl gebrauchter Nebensätze (in Relation zur Gesamtzahl finiter Verben), z.B. mit Hilfe des *Subordinationsindexes*, relative Anteile grammatischer Wortklassen (Konjunktionen, Präpositionen, Adjektive) u.dgl.m. (s.S. 41ff.) stellen hier relevante Themen dar. Methodisch wird man sich bei solchen Untersuchungen sehr stark an content-analytische Vorbilder (formaler Art) anlehnen müssen (vgl. Heller 1966, S. 300 ff.).

Auf allen drei Ebenen sprachlicher (psycholinguistischer) Analysen werden spezifische Gruppendifferenzen zwischen Hörschädigten und Normalhörenden wertvolle Informationen im Hinblick auf die Kognition Hörsprachbehinderter liefern. Darüberhinaus sind Aufschlüsse über intraindividuelle Veränderungen resp. lernzielbezogene Leistungskontrollen erwünscht und auf den angegebenen Wegen möglich. Die zuletzt berührte Thematik führt wieder in den Problembereich objektivierter *Schulleistungsmessung* (vgl. Gaude u. Teschner 1970).

Erkenntnisse über die Zusammenhänge von Sprache und Begabung (Intelligenzentwicklung) gewinnen für die *Sprachpädagogik* erhebliche Bedeutung, die vielleicht nirgendwo so deutlich zutage tritt wie in der Hörschädigtenpädagogik. Parallelen zwischen der historischen Taubstummepädagogik und der modernen Psycholinguistik bezüglich der Bewertung sprachlicher Voraussetzungen für die geistige Entwicklung des Menschen sind unverkennbar. Sowohl die (ältere) *sensorische* als auch die (jüngere) *soziokulturelle Deprivationshypothese* versuchen gleichartige Phänomene defizitären Sprachverhaltens zu erklären. Dabei wird die Homologie der Strukturen nicht nur unter Effizienzgesichtspunkten, sondern auch hinsichtlich formalätiologischer Kriterien (Deprivation) offenkundig. Eine Reihe empirischer Untersuchungsbefunde, etwa auffallende Parallelen im sprachlichen und Denkverhalten (fast sämtliche Merkmale des „restringierten“ Kodes sozio-kulturell deprivierter Kinder könnte man auch für taube und resthörige oder schwerhörige Kinder reklamieren, ähnliche Übereinstimmungen zeigen sich auf der kognitiven, insbesondere intellektuellen Verhaltensebene), weist auf die enge Beziehung von Milieu (Sensorium) und Sprache versus Sprache und Begabung hin. Ohne den Streit um die Determinismushypothese hier erneut aufzurollen (s.S. 43ff.), lassen sich u.E. aus den genannten Beobachtungen folgende Handlungskonsequenzen ableiten:

1. Die *sprachliche Förderung* Hörschädigter ist im Hinblick auf die kognitive, im besonderen intellektuelle Entwicklung von grundlegender Bedeutung. Im Zusammenhang damit übernimmt die *begabungsdiagnostische Aussage eine wichtige Hilfsfunktion* – sowohl bei der Ermittlung dispositioneller Voraussetzungen als auch bei der Kontrolle entsprechender Lernleistungen, die zugleich eine Begabungsförderung bedeuten. Allein die Tatsache, daß die Phänomene der Intelligenz im Kindes- und Jugendalter niemals starre Größen darstellen, erfordert regelmäßige Begabungsprüfungen während der Schulzeit.

2. Je früher die Bildungsbemühungen einsetzen, desto erfolgversprechender dürfte die individuelle Begabungsentwicklung in der Regel verlaufen. Dies gilt nach den vorliegenden Erfahrungen in besonderem Maße für sozio-kulturell benachteiligte Kinder (vgl. u.a. Nickel 1969). Im sonderpädagogischen Raum hatte man verhältnismäßig früh die Bedeutung einer gezielten *Vorschulerziehung* erkannt und beispielsweise in der Hörsprachgeschädigtenpädagogik zu realisieren versucht (vgl. Löwe 1962, 1964 u.a.). Naturgemäß rückt hierbei die



*Sprach(an)bildung* in den Vordergrund, deren Einfluß auf die intellektuelle Entwicklung Hörgeschädigter jedoch nicht unterschätzt werden darf. Daneben können Diskriminationsübungen, nonverbale Klassifikationen (Kategorisierungsleistungen bzw. nonverbales Begriffslernen), „Denkspiele“ u.ä. kognitive Prozesse der verschiedensten Art unterstützen (siehe hierzu Hudelmayer 1970, Kratzmeier 1970 u.a.).

3. Die *Forschung* im Bereich der Hörgeschädigtenpädagogik sollte *kooperativ* mit einschlägigen sozialwissenschaftlichen Disziplinen betrieben werden. Dabei wären stärker als bisher Erkenntnisse der modernen *Psycholinguistik* sowie der *Begabungs- und Bildungsforschung* zu berücksichtigen. Beispielsweise ist das Thema einer Adaptation des „restringierten“ Gebärdenkodes an den „elaborierten“ Laut/Schriftsprachkode bislang fast ausschließlich unter sprachwissenschaftlichen bzw. sprachdidaktischen Aspekten diskutiert worden. Auf die sozial- und persönlichkeitspsychologischen oder bildungssoziologischen Probleme wurde unseres Wissens kaum eingegangen. So besteht der Verdacht, daß für den Gehörlosen der Übergang von der (schulischerseits tabuierten) Gebärdensprache hin zur (schulischerseits erwünschten resp. bekräftigten) Phonemsprache nicht weniger einem Subkulturwechsel gleichkommt als für das Unterschichtkind die Adaptation des öffentlichen (restringierten) an den elaborierten Sprachkode. Mögen auch die angesprochenen Schwierigkeiten beim hörgeschädigten und bildungsmäßig unterprivilegierten (vollsinigen) Kinde graduell unterschiedlich sein, im Kern ergibt sich doch die gleiche Problematik einer „Mentalitätssperre“, die es *psychologisch* aufzulösen gilt (siehe auch S. 47ff.). Rigorose Unterdrückung unerwünschter Gebärden(-Sprachansätze) aus sprachlogischer oder unterrichtspragmatischen Gründen erscheint im Lichte psycholinguistischer Untersuchungsbefunde sowie im Hinblick auf die praktischen Ergebnisse solchen Tuns mehr als zweifelhaft. Daß über das „Wie“ der Verwirklichung obiger Ziele bislang auf psycholinguistischer/bildungssoziologischer versus gehörlosenpädagogischer Seite weithin noch Unklarheit vorherrscht, unterstreicht einmal mehr die Notwendigkeit zur interdisziplinären Zusammenarbeit.

## II. Einführung in die Testtheorie

Auf der Erde existieren zur Zeit schätzungsweise 10 000 standardisierte psychologische Tests, von denen mindestens 30% Intelligenzmeßverfahren darstellen. Als standardisierter Test wird ein Untersuchungsverfahren angesprochen, das bezüglich seiner Gültigkeit (*Validität*), Meßgenauigkeit (*Reliabilität* oder Zuverlässigkeit) und *Objektivität* (sowohl der Testdurchführung als auch der Ergebnisauswertung) eingehend kontrolliert worden ist. Darüberhinaus müssen in der sog. *Itemanalyse* (Aufgabenanalyse) Schwierigkeit und Trennschärfe jedes einzelnen Testitems bestimmt sowie im Hinblick auf die Klassifizierung der Testleistung der untersuchten Personen (*Probanden*) spezifische *Testnormen* erstellt worden sein, bevor man das Prädikat *standardisierter* Test einem diagnostischen Untersuchungsverfahren zusprechen sollte. Um die Befunde im empirischen Teil dieses Werkes würdigen zu können sowie im Hinblick auf die psychodiagnostische Praxis erscheint eine genauere Kenntnis testtheoretischer Grundlagen unerlässlich. Die folgenden Ausführungen verstehen sich deshalb als knappe Einführung in die klassische Testtheorie, wobei sich d. Verf. an einen früheren Aufsatz (vgl. Heller 1965) anlehnt.

Wenn wir uns auf die Darlegung der *klassischen* Testtheorie beschränken, dann sind wir hier eine kurze Erklärung schuldig. Zwar werden allerjüngstens einige Grundannahmen der klassischen Testtheorie einer kritischen Überprüfung unterzogen (siehe Fischer 1968), doch wäre es u.E. verfrüht, hieraus eine „Verdammung“ der klassischen Testtheoreme abzuleiten. Stochastische resp. logistische Testmodelle, wie sie in der genannten Veröffentlichung von Fischer beschrieben werden – am bekanntesten ist wohl Raschs kreative Theorie psychologischer Messung –, mögen eines Tages die Stelle bisher gültiger Theorien einnehmen, vorerst freilich wäre es töricht, einigermaßen gesicherte Grundlagen zugunsten problematischer und vor allem noch wenig praktikabler Modellansätze aufzugeben. Andererseits sollte man die Entwicklung auf diesem Gebiet sehr sorgfältig beobachten und sich möglichen Verbesserungen oder einer Theorieerweiterung, ggf. auch radikalen Veränderungen, zu einem gesicherteren Entwicklungsstadium (später) nicht entgegenstellen. Vorerst freilich wird man sich an die weithin anerkannten „klassischen“ Grundlagen der Testtheorie halten müssen.

### 1. Historische Aspekte

Das Wort „Test“ (lat. testimonium; engl. test) bedeutet ursprünglich *Prüfung*, Stichprobe, Zeugnis, Beweis. Als Vorläufer heutiger Testprüfungen können die Initiationsriten vieler Primitivgesellschaften, ihre Mut- und Tüchtigkeitsproben (zur Ermittlung der Eignung des Heranwachsenden für die Übernahme der Erwachsenenrolle) oder „Begabungsprüfungen“ in Form von Verstandesfragen, Rätselraten (vgl. Mythen wie Ödipus und die Sphinx oder Turandot) gelten. Weitere Beispiele prähistorischer Testmethoden finden sich bei Platon, der im Dialog vom Staat ein Auswahlverfahren für die Wächter (Krieger) des idealen Gemeinwesens beschreibt, wobei Mut, Selbstdisziplin, Unbestechlichkeit u.ä. Eigenschaften erprobt werden (nach Hofstätter 1957, S. 287 f.). Auslese- und Eignungsprüfungen zur Erfassung körperlicher versus intellektueller Tüchtigkeit wurden fast in allen Kulturen und zu allen Zeiten praktiziert; die Berichte von den Griechen, Römern, Juden bis hin zu den modernen Beispielen der Heerespsychologie, der psychologischen Dienste in Schule und Berufsberatung liefern dafür genügend Belege. Interessanterweise verwandte schon Sokrates in Athen ein „intellektuelles“ Prüfsystem, das im Prinzip heutigen Vorstellungen über Lernleistungstests (sog. Informelle oder Kriterientests bzw. Teacher-made Tests) sehr nahekommt, nämlich eine Kombination aus Lehren (Lernen) und Prüfen; siehe dazu besonders Gaude u. Teschner 1970.

Als der eigentliche Vater des Intelligenztests wird jedoch vielfach Sir Francis Galton, ein vermutlicher Neffe Ch. Darwins, angenommen. Er interessierte sich wie kaum jemand

vor ihm für die *interindividuellen Differenzen* menschlicher Eigenschaften und Fähigkeiten und sprach bereits um die Mitte des vorigen Jahrhunderts vom „Test als Kennwort einer Methode“.<sup>34</sup>

Im Jahre 1890 veröffentlichte dann der Wundt-Schüler J. McKean Cattell seinen vielzitierten Aufsatz „Mental tests and measurement“. Beide Forscher verwandten nicht nur erstmals in der Literatur den Begriff ‚Test‘ im modernen Sinnverständnis, ihre grundlegenden Gedanken zur psychologischen Diagnostik sind bis auf den heutigen Tag aktuell. Sicherlich erwarben sich auch andere Forscher historische Verdienste um die Entwicklung der Testdiagnostik – besonders erwähnenswert sind hier die französischen Psychiater Itard und Seguin mit ihrem „Formenbrett“ zur Prüfung der (Handlungs-)Intelligenz Schwachsinniger sowie St. Hall mit seinem Fragebogen (Kenntnisprüfung) und der Würzburger Psychiater Rieger, der kasuistisch arbeitete. Als Vertreter der Pädagogik bzw. als pädagogisch interessierte Psychologen traten besonders hervor Ebbinghaus (1885, 1897) mit seinen Untersuchungen über das Gedächtnis und seinem berühmt gewordenen „Lückentest“ (zur Erfassung intellektueller Kombinationsleistungen<sup>35</sup>), ferner Meumann (1905), Binet (1898, 1905 ff.) bzw. Binet und Mitarbeiter, besonders Henri und Simon, in Deutschland wieder Stern, Bobertag und Hylla, um nur wenige Namen zu nennen.

Das herausragende Verdienst Galtons um den differentialdiagnostischen Fortschritt in bezug auf die Intelligenzerfassung ist zweifellos der Tatsache zuzuschreiben, daß er schon die Bedeutung *statistischer* Erhebungen und Vergleiche erkannt und für die Testentwicklung nutzbar gemacht hat. Galtons Test sollte Tätigkeiten, Fertigkeiten und Leistungen in quantitativen Resultaten festhalten. Im Gegensatz zur Leipziger Schule lag ihm nichts an der Aufdeckung der psycho-physiologischen Bedingungen. Noch in der heutigen Testpsychologie ist dieser „behavioristische“ Zug nachweisbar. Galtons Testaufgaben umfaßten physische und psychische Messungen nacheinander. Für ihn, den Anthropologen, bedeutete die Psychometrie lediglich einen Sonderfall der Anthropometrie – ähnlich wie auch bei Alfred Binet später.

Indem Galton so dem Prinzip der *biologisch-statistischen Messung* Geltung verschuf, führte er noch eine weitere wichtige Erkenntnis in die psychologische Wissenschaft ein. Der belgische Astronom Quetelet hatte bereits die Gültigkeit des „Gesetzes der Abweichungen vom Durchschnitt“ (ursprünglich von Laplace und Gauß erarbeitet) auch auf biologische Messungen ausgedehnt. Hiermit konnte nun Galton demonstrieren, „daß nicht nur in absoluten Maßeinheiten, sondern auch durch die Feststellung der *relativen Position in einer Gruppe* in Form der Abweichung vom Mittelwert gemessen werden konnte“<sup>36</sup>.

Der entscheidende Gedankenschritt war hierbei die Idee Galtons, daß die geistigen Fähigkeiten analog den biologischen Eigenschaften (Gewicht, Größe, Kopfumfang u.a. Körperindizes) verteilt sein müßten, „Intelligenz sah er als eine biologische Eigenschaft, formulierte ihre Normalverteilung als Hypothese und demonstrierte die Hypothese empirisch durch das Studium der Verteilung von Zensuren der Absolventen des Royal Military College in Sandhurst<sup>36</sup>“. Interessanterweise kritisierte in diesem Zusammenhang Galton bereits die Unterschiedlichkeit der Bewertungsskalen (Lehrerurteile) und forderte einen einheitlichen Maß-

---

<sup>34</sup>Zur Information bezüglich der in diesem Abschnitt erwähnten Literaturquellen, die nicht in unserem Buchregister aufgeführt werden konnten, verweisen wir auf den lesenswerten Artikel von Groffmann (1964), worin die historische Entwicklung der Intelligenzmessung eingehend dargestellt wurde.

<sup>35</sup>Die Kombinationsmethode nach Ebbinghaus (loc. cit., S. 433 f.) „bildet ein einfaches, zu seiner Anwendung nur wenig Zeit forderndes Prüfungsmittel für die in Schule und Leben wahrhaft wichtige und wertvolle Betätigung des Geistes, sie ermöglicht eine eigentliche Intelligenzprüfung“ (zit. bei Groffmann 1964, S. 160).

<sup>36</sup> Zit. nach Groffmann 1964, S. 155; Kursiv v. Verf.

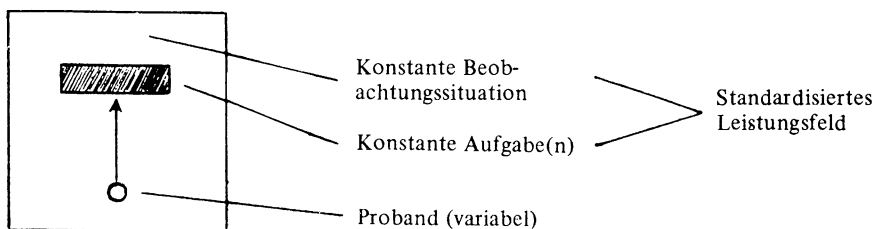
stab, den er mit Hilfe der Gesetze der *Normalverteilung* und der *Abweichung vom Mittelwert* vorschlug. Eine Klassifizierung verschiedener Intelligenzstufen durch Tests war somit via Normalverteilungshypothese ermöglicht. 1869 schrieb Galton: „So kommen wir denn zu der unausweichlichen, wenn auch unerwarteten Schlußfolgerung, daß hervorragend begabte Menschen sich soweit über die Mittelmäßigkeit erheben, wie die Idioten darunter gedrückt sind<sup>37</sup>“. Sicher ahnten diese Zusammenhänge erfahrene Lehrer und Ärzte auch schon früher, aber Galton formulierte sie erstmals so prägnant und überzeugend.

## 2. Grundstruktur, Gegenstandsdefinition und Phasen des testdiagnostischen Prozesses

Bevor wir die Grundstruktur testdiagnostischer Aussagengewinnung skizzieren, knüpfen wir noch einmal an die Versuche zur Erfassung der *individuellen Differenzen* an. Einen ersten experimentellen Beitrag zum Problem individueller Differenzen lieferte der Astronom Bessel mit seinen 1876 veröffentlichten Untersuchungen zur Frage der „persönlichen Gleichung“.<sup>38</sup>

In Wundts Leipziger Laboratorium, dem ersten (eigenständigen) Psychologischen Institut überhaupt, konnte noch zu Ende des letzten Jahrhunderts nachgewiesen werden, daß seelische Vorgänge experimenteller Forschung zugänglich sind. Von diesen frühen experimentalpsychologischen Arbeiten ging ein nicht geringer Schulungseffekt auf die spätere Testentwicklung aus. Methodenproblem Nr. 1 war dabei die genaue Kontrolle und Beherrschung der in die Experimente eingehenden Bedingungen. So erkannte man bereits, daß die äußeren Bedingungen, etwa Unterschiede in der Reizdarbietung oder variierte Versuchsanordnungen, die Versuchsergebnisse nicht unwesentlich beeinflussten, wodurch das Standardisierungsproblem deutlicher ins Bewußtsein rückte. Weit schwieriger freilich gestaltete sich die Kontrolle der „inneren“ Versuchsbedingungen, d.h. individueller Reaktionsunterschiede (der Versuchspersonen selber), womit der eigentliche *Gegenstand der Diagnostik* benannt wäre: *die interindividuellen Differenzen*.

Als einer der ersten entwickelte J. McKean Cattell ein Testmodell, an dem sich die *Grundstruktur des Diagnostizierens* veranschaulichen läßt: Im determinierten (standardisierten) Leistungsfeld mit *konstanter Beobachtungssituation* und *konstanter Aufgabenstellung* erscheint als *einzige Variable der Proband*. Diese Grundstruktur ist für alle testdiagnostischen Prozesse relevant.



<sup>37</sup> Nach Groffmann zit.

<sup>38</sup> Nach Drever/Fröhlich (1968, S. 170) steht der Ausdruck „persönliche Gleichung“ (personal equation) als „Bezeichnung für die von dem Astronomen Bessel (1876) erstmals berichtete Tatsache, daß zwei gleichartig geschulte Beobachter in ihren Reaktionszeiten voneinander abweichen, wobei jeder einzelne Beobachter relativ gleichartige Fehlergrößen aufweist, die zeitlich überdauernd sind... Die Bezeichnung „persönliche Gleichung“ resultiert aus der Annahme der relativen Konstanz der individuellen (persönlichen) Fehlergröße.“

Der Hauptvorteil dieses Testmodells liegt darin, daß man unter den genannten Standardbedingungen – *gleicher* Beobachtungssituation und Aufgabenstellung für alle Probanden – die individuellen Reaktionsweisen (der Probanden) isolierend miteinander *vergleichen*, d.h. interindividuelle Leistungsunterschiede *messen* kann. Die Systematisierung testdiagnostischer Prozesse ermöglicht somit eine *objektivere* Urteilsfindung, z.B. über die Intelligenz, die Schuleignung, die Schulleistung u.ä., als die unsystematische Gelegenheitsbeobachtung, etwa im Unterricht, in der Freizeit, in der Familie oder im Internat. Unter Standardbedingungen gewonnene Prüfergebnisse oder Testergebnisse lassen sich nicht nur leichter quantifizieren, sie gestatten auch objektive und zuverlässige Rückschlüsse auf die den *Testleistungen* zugrundeliegenden (mehr oder weniger hypothetisch postulierten) Persönlichkeitsmerkmale, Fähigkeiten u.dgl.m.

Historisch gesehen gelangte man also erst auf dem Umwege über die Methodenkontrolle in der Experimentalforschung zum Phänomen der (inter)individuellen Differenzen, an denen die „reinen“ Psychologen wie Wundt und Mitarbeiter sachlich überhaupt nicht interessiert waren; vielmehr versuchte man zuerst, diese „Störvariablen“ im Experiment auszuschalten – ohne Erfolg! Ihre Bedeutung als eigenen Forschungsgegenstand der Testdiagnostik erkannten als erste Wundts Schüler McK. Cattell, Krapelin und Münsterberg.

Während die Experimentalpsychologie auf die Erforschung allgemeingültiger „Gesetze“ – allgemeinpsychischer Beschreibungsmodi resp. Bedingungen des Verhaltens – aus war (ihr Hauptgegenstand sind deshalb die Inhalte der „Allgemeinen Psychologie“), stand bei der Psychologischen Diagnostik von vornherein das individuell Eigenartige im Vordergrund des Interesses. Diese Differenzierung kommt auch in W. Sterns (1911) Unterscheidung zwischen „Forschungsexperiment“ (auf das Allgemeine, Gesetzmäßige gerichtet) und „Prüfexperiment“ (auf das Spezielle, individuell Eigenartige gerichtet = diagnostisches Experiment) zum Ausdruck. Nach ihm ist der *Test* dasjenige Prüfungsexperiment, „das bestimmt ist, in einem gegebenen Fall die individuelle psychische Beschaffenheit einer Persönlichkeit oder eine einzelne psychische Eigenschaft an ihr festzustellen“.

Im Kern identisch, jedoch operationaler definiert L. Michel (1964, S. 19): „Ein psychodiagnostischer Test kann als spezifisches psychologisches Experiment gekennzeichnet werden, das der Erkundung spezifischer individueller Unterschiede dient. Es besteht im Wesentlichen darin, daß unter standardisierten Bedingungen eine Verhaltensstichprobe des Probanden (Pbn) provoziert wird, die einen wissenschaftlich begründeten Rückschluß auf die individuelle Ausprägung eines oder mehrerer psychischer Merkmale des Pbn gestattet.“ So können wir zusammenfassend festhalten:

*Ein psychologischer Test soll interindividuelle Differenzen messen, d.h. die Position eines Individuums innerhalb seiner Bezugsgruppe (Alters- oder Schul-/Klassengruppe, Geschlechts- bzw. Berufsgruppe bzw. einer beliebigen Referenzpopulation) hinsichtlich bestimmter Merkmale oder Fähigkeiten (quantitativ) bestimmen.* Im Prinzip gilt diese Definition für alle Tests, für die psychometrischen (Intelligenztests, Leistungstests u.ä.) wie die sog. Projektiven Tests (z.B. Rorschach-Test, Thematic Apperception Test).

Nach dieser Begriffsbestimmung können wir uns nun den *Hauptphasen des testdiagnostischen Prozesses* zuwenden. Michel (a.a.O., S. 20 f.) schlägt eine Gliederung in vier Phasen vor, wobei natürlich zu bedenken ist, daß jede Phaseneinteilung – auch diese – mehr oder weniger eine Abstraktionsleistung repräsentiert, d.h. konkret diese scharfe Stufentrennung nicht immer möglich sein wird.

1. Das Verhalten des Pbn wird durch eine standardisierte Reizfiguration provoziert. Im Idealfalle erfolgt die *Provokation* des Verhaltens unter konstanten Außenbedingungen (der unabhängigen Variablen), wobei einzig das Verhalten der Pbn selbst (abhängige Variable) variiert (vgl. Schema auf Seite 57).

2. Im nächsten Schritt muß das provozierte Verhalten des Pbn beobachtet und beschrieben (in der Regel schriftlich fixiert) werden. Dabei wird die **Registrierung** des Testverhaltens jeweils selektiv und zentriert auf diagnostisch bedeutsame Verhaltensdaten erfolgen, im Intelligenztest beispielsweise auf **intellektuelle** Verhaltensweisen und nicht bzw. weniger auf soziale oder persönlichkeitspsychologische Dimensionen (Extraversus Introversion, Ängstlichkeit, Neurotizismus u.ä.). Diese müßten u.U. durch andere Verfahren (Zusatztests, Fragebogen) bzw. Exploration und Verhaltensbeobachtung (Observational Techniques) erfaßt werden. Stets ist jedoch darauf zu achten, daß die Registrierung des Verhaltens **objektiv** geschieht; präzise Hinweise des Testautors, z.B. genau festgelegte Regeln, müssen sicherstellen, daß verschiedene Testleiter ein- und dasselbe Testverhalten identisch registrieren.

3. An die Protokollierung schließt sich die **Auswertung** der registrierten Verhaltensdaten an. Bei Intelligenztests bzw. den psychometrischen Testverfahren werden in der Regel die richtigen Lösungen ausgezählt und zu Teilscores (Summe der Rohpunkte in den Untertests) oder/und Gesamtscorewerten (Rohpunktschsumme insgesamt) addiert. Für den Vergleich der individuellen Scorewerte mit den entsprechenden Gruppennormen ist gewöhnlich eine Transformation der Rohwerte in Standardwerte bzw. Prozentränge notwendig; die im Anhang beigefügte Transformationstabelle gängiger Testnormen gestattet eine rasche Orientierung über die Position eines Individuums (Pb) innerhalb der Referenzpopulation bei Testergebnissen unterschiedlicher Provenienz resp. unter Benutzung unterschiedlicher Normenskalen (vgl. Seite 245, Tab. VIII). Außer der Ermittlung numerischer Kennwerte erfolgen mitunter weitere Verarbeitungen zu Syndromen, graphische Profildarstellungen u.dgl.m., um die Durchschaubarkeit des Datenmaterials zu verbessern.

4. In der **Interpretationsphase** werden nun die diagnostischen Schlußfolgerungen aus dem aufbereiteten Material gezogen. Dabei wird angenommen, daß ein logischer bzw. empirisch nachgewiesener Zusammenhang zwischen den beobachteten Testresultaten und den postulierten psychischen Verhaltensdimensionen besteht (siehe hierzu das Validitätsproblem auf Seite 72ff.). Ähnlich wie bei den vorherigen Phasen muß auch für die Interpretation wieder **Objektivität** gefordert werden, d.h. aus identischen Testresultaten müssen stets – auch von verschiedenen Experten (Lehrern, Psychologen, Ärzten versus Lehrer A, Lehrer B usw.) – gleiche Schlußfolgerungen gezogen werden.

### 3. Psychometrische Prinzipien und Konzepte (intelligenz)diagnostischer Untersuchungsverfahren

#### a) Allgemeine Probleme der Standardisierung

**Standardisierung** und damit **Kontrollierbarkeit** der Untersuchungsbedingungen sind Grundforderungen an einen modernen Test. Obwohl diese Probleme sehr eng mit den noch zu erörternden Testgütekriterien der Objektivität und Reliabilität in Zusammenhang stehen, sollen einige „äußere“ versus „innere“ Testbedingungen bzw. soziale Situationsvariablen vorweg angesprochen werden. Wie das Schaubild auf Seite 57 bereits verdeutlichte, gehören zu den wichtigsten Standardbedingungen diagnostischer (testpsychologischer) Untersuchungen **konstante Aufgabenstellung** (gleiche Testmaterialien für alle Pbn) und **konstante Versuchsbedingungen** (gleiche Instruktion resp. Administration des Tests). Selbst geringfügige und unerfahrenen Testern oft als nebensächlich erscheinende Variationen in der Testdarbietung können bereits die Testresultate beeinflussen. Es empfiehlt sich deshalb immer, sich möglichst streng an die vorgeschriebenen Untersuchungsmodalitäten zu halten; mehr oder weniger subjektiv empfundene Mängel der Testinstruktion oder Testdarbietung sollten – trotz

vielleicht berechtigter Kritik – vom Testleiter nicht eigenwillig „verbessert“, d.h. abgeändert werden, da manipulierte Standardprüfbedingungen in der Regel eine weit größere Gefahr für die Objektivität der Untersuchung darstellen als vermeintliche oder wirkliche Unebenheiten in der Testanordnung. Bei offensichtlichen Fehlern empfiehlt es sich, Testautor und Verlag, gegebenenfalls auch potentielle Testbenutzer über einschlägige Publikationsorgane, umgehend zu informieren. Freilich garantiert eine strikte Befolgung der gedruckten Testanweisung noch nicht, daß damit auch schon notwendig für alle Pbn Konstanz, also unter Standardbedingungen vergleichbare, Voraussetzungen gegeben sind. Dazu bedarf es weiterer Kontrollen der Testsituation i.w.S.

Zu den *äußeren* Bedingungen wären u.a. Wetterlage, Art und Ausstattung des Testraumes, Schreibunterlage und sonstige notwendige Utensilien, Zeitpunkt resp. Tageszeit der Untersuchung u.dgl.m. zu rechnen. Bereits ein abgebrochener Bleistift kann (besonders bei Speedtests) als Störfaktor auftreten. Der Testleiter sollte deshalb für einen störfreien Untersuchungsablauf Sorge tragen. Ein Hinweisschild an der Tür des Testraumes („Bitte nicht stören, Testuntersuchung!“), vorrätige Bleistifte, Radiergummis, Schreibpapier usw. müssen eingeplant, voraussehbare Störungen (z.B. Pausenlärm) vermieden werden. Ein guter Testleiter wird sich – genauso wie ein guter Lehrer bei der Durchführung einer Klassenarbeit – jeweils um *optimale* „äußere“ Bedingungen für seine Schüler kümmern.

Wer seine Testuntersuchungen sorgfältig vorbereitet, wird auch dafür sorgen, daß diese während der „normalen“ Schulzeit, in der Regel also vormittags, durchgeführt werden. Auf jeden Fall verbieten sich Intelligenz-/Leistungstests am Spätnachmittag oder Abend, wenn die Leistungskurve der (meisten) Schüler rapide absinkt. Überhaupt gilt die Forderung, daß sich der Proband während der Testprüfung psychophysisch in „normaler“ (guter) Verfassung befinden solle. Im einzelnen wären als *innerseelische* Bedingungen zu nennen:

Einfluß *vorhergehender Tätigkeiten*, wobei sich ein „schönes“ Erlebnis oder eine interessante, in der vorangegangenen Unterrichtsstunde behandelte Thematik im allgemeinen leistungssteigernd für die (nachfolgende) Testuntersuchung auswirkt. Dieser Zusammenhang besteht besonders bei jüngeren Kindern;

Abhängigkeit der Testleistung von der *Motivation* des Pb, wobei sich emotionale Belastung, Affekte, Entmutigung, Gleichgültigkeit und Desinteresse nachteilig versus Anreize wie Belohnungsversprechen, Lob oder (seltener) Tadel vorteilhaft auf das Testleistungsverhalten auswirken. Ein in diesem Zusammenhang bedeutsames Problem stellt die *Testangst* dar. Für die Mehrzahl der Pbn besteht eine negative Korrelation zwischen Testangst und Testleistung (je mehr Examenscharakter, desto größer die Angst und desto niedriger die Leistung!), doch lassen sich die angesprochenen Verhältnisse keineswegs für alle Menschen auf diese Weise aufklären. In Kontrolluntersuchungen wurden auch gegenläufige Reaktionen – besonders bei Mädchen – beobachtet, was freilich nicht von der allgemeinpädagogischen Forderung nach Angstvermeidung enthebt. Ferner konnte nachgewiesen werden, daß *Argwohn* und *Mißtrauen* (gegenüber der Testung) seitens der Pbn die Leistung in der Testprüfung beeinträchtigen. Häufig lassen sich solche Motivationsbarrieren (vornehmlich bei älteren Kindern bzw. Jugendlichen) dadurch beseitigen, daß man eine wahrheitsgetreue Zielbeschreibung der Testuntersuchung gibt resp. verspricht, die einzelnen Schüler auf Wunsch individuell über das Testergebnis zu informieren. Überhaupt empfiehlt sich vor jeder Untersuchung ein offenes Wort über Sinn und Zweck der Testung – natürlich in kindgemäßer Diktion und ohne Dramatisierung der Prüfsituation;

*Testtraining* und „*Betrügen*“ (Mogeln) können ebenfalls die Resultate verfälschen. Beide Verhaltensweisen resultieren aus der Absicht, sich Vorteile gegenüber den Testbedingungen (Erfolgchancen) der Mitkonkurrenten zu verschaffen und treten besonders dann auf, wenn das Testergebnis für den Pbn von ausschlaggebender Bedeutung ist oder erscheint. Der Gefahr des Mogels versucht man durch die Erstellung von Paralleltestformen (mit jeweils anderen, aber gleichschweren Aufgaben) oder den Einbau von sog. Lügenfragen (z.B. bei Persönlichkeitsfragebogen) zu begegnen. Die Gefahr der Ergebnis-

fälschung durch vorheriges Aufgabentraining ist bei allen bekannten Intelligenztests und in besonderem Maße bei Schulleistungstests gegeben. Tests müssen demnach verschlossen aufbewahrt werden. Ferner sollte man aus Gründen des Testschutzes – trotz „pädagogischer Verlockungen“ – im einzelnen keine Testergebnisse (besonders keine Falschlösungen) mit den Pbn. resp. Schülern besprechen. Auch bei *Testwiederholungen* ist ein mehr oder weniger großer Leistungszuwachs (Lerngewinn) in Rechnung zu stellen, nach unserer Erfahrung mit Retestungen beispielsweise beim Progressiven Matrizentest von Raven (innerhalb weniger Wochen oder Monate) durchschnittlich 10 bis 20 %. Freilich variieren die individuellen Zuwachsquoten z.T. beträchtlich – selbst bei relativer Konsistenz intraindividuelle Merkmalsausprägung. Bei wiederholten Untersuchungen wird man deshalb auf die Parallelfom(en) desselben Testverfahrens zurückgreifen oder/und äquivalente Testverfahren alternierend einsetzen, um den Abnutzungseffekt bekannter Tests möglichst klein zu halten.

Schließlich sei noch die *soziale Situation* als Einflußvariable der Testleistung angeführt. Hierher gehören der Bekanntheits- versus Fremdheitsgrad des Testleiters und der getesteten Gruppenmitglieder genau so wie der Einfluß emotionaler Komponenten im Interaktionsfluß zwischen Testleiter und Proband („warm“ versus „kalt“), vertraute Unterrichts- oder Erziehungsstile, Sympathie versus Antipathie, Geschlechtsvariable u. dgl.m.

Der Möglichkeit zur *Standardisierung der psychosozialen Testsituation* wie überhaupt der diagnostischen Situation sind also realiter Grenzen gesetzt. Vorab dem Testleiter obliegt es, potentielle Störeinflüsse zu vermeiden bzw. aktuelle Determinanten festzuhalten und spätestens bei der Interpretation der Testresultate zu berücksichtigen. Sofern entsprechende Kontrollskalen nicht bereits in Tests eingebaut (vgl. die Lügenliste im MMPI) oder als Zusatzverfahren (z.B. die Testangstskale von Sarason, der Kinder-Angst-Test von Thurner und Tewes u.ä.) verfügbar sind, ist der Testleiter hierbei ausschließlich auf eine sorgfältige Verhaltensbeobachtung angewiesen. Ausführlicher gehen L. Michel (1964, S. 22 ff.) sowie H. Hartmann (1970, bes. S. 34-69) auf die skizzierten Standardisierungsprobleme ein. Siehe ferner Cronbach & Gleser 1965, Sader 1957, Sader & Keil 1966, Spitznagel 1964 u.a.

## b) Die Aufgabenanalyse (Itemanalyse)

Die Güte eines Tests ist abhängig von der Qualität seiner Aufgaben (Items). Diese wird in der sog. Itemanalyse ermittelt, die somit die erste wichtige Etappe im Testaufbau resp. der Testanalyse darstellt. Die Güte eines Testitems bestimmt sich vorab durch seine *Schwierigkeit* und seine *Trennschärfe*, womit schon die Hauptziele der Itemanalyse benannt sind. Diese erfolgt auf empirischem Wege, indem die Testvorform (ein von Experten – bei Intelligenz- bzw. Schultests in der Regel Psychologen und Pädagogen – zusammengestelltes Inventar geeigneter Aufgaben) einer möglichst repräsentativen Analysenstichprobe dargeboten wird. Damit verfolgt man ein zweifaches Ziel: „1. Revision des Testes im Hinblick auf eine höhere Reliabilität und Validität durch Ausschaltung und Verbesserung nicht genügend geeigneter Aufgaben. 2. Überprüfung der Testpunkteverteilung hinsichtlich einer höheren Reliabilität und einer besseren Normierbarkeit“ (Lienert 1967, S. 70). Mit den Testgütekriterien (Reliabilität und Validität) werden wir uns gleich noch eingehender zu beschäftigen haben. Hier interessieren zunächst die Methoden zur Aufgabenverbesserung, d.h. die Ermittlung der Schwierigkeit und Trennschärfe jeder einzelnen Testaufgabe.

Der Grad der Schwierigkeit einer Testaufgabe wird in einem statistischen Kennwert, dem sog. *Schwierigkeitsindex* (P) ausgedrückt. Dieser meint den Prozentsatz der auf eine bestimmte Testaufgabe entfallenen Richtig-Antworten. Seine Definitionsformel lautet demnach:  $P = 100 \cdot N_R / N$ . Hierbei bedeutet  $N_R$  = Anzahl der Pbn, die die betr.



Aufgabe richtig gelöst haben;  $N$  = Anzahl der Pbn, die die betr. Aufgabe bearbeitet haben (Stichproben- $N$ ). Es ist zu beachten, daß der Schwierigkeitsindex (Wert  $P$ ) bei schweren Aufgaben „niedrig“ und bei leichten Aufgaben „hoch“ ausfällt, weshalb man unter dem Gesichtspunkt *psychologischer* Interpretation besser von einem „Leichtigkeits“-Index sprechen würde. Doch hat sich in der Fachliteratur der erstgenannte (statistische) Begriff durchgesetzt, woran auch wir festhalten wollen. Das Begriffspaar *Schwierigkeit* versus *Leichtigkeit* (ohne Spezifikation) soll hingegen im psychologischen Sinnverständnis gebraucht werden. Ausführlicher haben wir die Prozedur der Itemanalyse am Beispielfall der HHBT-Standardisierung beschrieben (vgl. Löwe und Heller 1972, S. 37 ff.).

Mit der *Trennschärfe* ist die Eigenschaft einer Testaufgabe angesprochen, die „guten“ von den „schlechten“ Pbn oder – im konkreten Anwendungsfalle – die überdurchschnittlich intelligenten von den unterdurchschnittlich intelligenten Pbn zu diskriminieren. Zur Bestimmung der Diskriminationsfähigkeit eines Testes und damit seiner Itemtrennschärfe bedient man sich einer Reihe von Methoden, deren präziseste die sog. „Item-Rohwertkorrelation“ (Vukovich 1964) darstellt. Dabei werden die Richtig-versus Falsch-Antworten des analysierten Testitems in Beziehung zum Gesamtpunktwert (Rohwert) des Testes gebracht. Dies geschieht mit Hilfe der sog. *punktbiserialen Korrelation*, deren Ergebnis (im Kennwert  $pbis^{\text{fit}}$  ausgedrückt) analog zum Produkt-Moment-Korrelationskoeffizienten interpretiert wird. Die Formel zur Berechnung des *Trennschärfekoeffizienten* (vgl. Lienert a.a.O., S. 94 f.) lautet:

$$pbis^{\text{fit}} = \frac{\bar{X}_R - \bar{X}}{s_x} \cdot \sqrt{\frac{p}{q}}, \quad \text{wobei } p = N_R/N \text{ und } q = 1 - p \text{ ist.}^{39}$$

Im allgemeinen ist man bemüht, optimale Schwierigkeitsindizes um 50 (zwischen 25 % und 75 % Richtig-Antworten) zu erhalten; diese Faustregel gilt zumindest für die Konstruktion der hier einschlägigen Intelligenz- und Leistungstests. Die Trennschärfekoeffizienten sollten im allgemeinen über 0.30 liegen (bei niedrigeren Werten muß die betr. Testaufgabe ausgeschieden oder verbessert werden); Koeffizienten um 0.4 bedeuten bereits recht gute, solche um 0.5 oder darüber (selten) sehr gute Trennschärfe der betr. Testaufgabe.

Mit diesen knappen Informationen wollen wir es hier bewenden lassen. Im empirischen Teil dieses Buches werden wir gelegentlich die angesprochene Thematik erneut aufgreifen, wenngleich dort andere Fragestellungen in den Vordergrund der Untersuchungsvorhaben rücken; Probleme der Itemanalyse werden vor allem im Zuge der Neukonstruktion von Tests aktuell. Im übrigen sei auf die sehr detaillierten Darstellungen bei Lienert (1967, S. 62 ff.) sowie Chauncey & Dobbin (1970, S. 147 ff.) verwiesen. Dort werden weitere Analysemethoden (Kurzmethoden) besprochen, die besonders für schulische Erfordernisse bzw. im Rahmen der Bearbeitung sog. Informeller oder Tea-

<sup>39</sup> Die Symbole in der angeführten Formel bedeuten:

$pbis^{\text{fit}}$	=	punktbiserialer Korrelationskoeffizient des/der Items,
$\bar{X}$	=	arithmetisches Mittel aller Testrohwerte,
$\bar{X}_R$	=	arithmetisches Mittel der Testrohwerte von jenen Pbn, die die betr. Aufgabe richtig beantwortet haben,
$s_x$	=	Standardabweichung der Testrohwerte aller Pbn,
$N$	=	Anzahl aller Pbn,
$N_R$	=	Anzahl derjenigen Pbn, die die betr. Aufgabe richtig beantwortet haben.

cher-made-Tests, auch Kriteriumtests genannt, interessant sein dürften (vgl. noch Gaude & Teschner 1970, S. 105 ff.; Ingenkamp 1971, S. 43 ff.; Wendeler 1969, S. 33 ff.).

### c) Testgütekriterien als Merkmale der Brauchbarkeit diagnostischer Untersuchungsverfahren

Über die Qualität und damit die Brauchbarkeit eines Tests geben die sog. Testgütekriterien Aufschluß: *Objektivität*, *Reliabilität* und *Validität*. Die Erfüllung dieser Qualitätsansprüche ist eine unabdingbare Forderung an jedes vollstandardisierte Meßverfahren, also auch an Intelligenztests – Verfahren, die den genannten Gütekriterien nicht oder nur sehr mangelhaft genügen, sollte man besser *nicht* als „Test“ ansprechen.

Je nach Untersuchungsziel, Methodenangebot und Anwendungspopulation wird man darüberhinaus praktische Erfordernisse der rationellen Durchführung bzw. Auswertbarkeit (z.B. Individual- oder Gruppentest, Schablonen-/Hand- oder automatische (elektronische) Belegauswertung), Fragen der Ökonomie (Arbeits-, Zeit- und finanziellen Aufwand) sowie Gesichtspunkte der Kinder- und Jugendpsychologie, der Sozialpsychologie u.ä. (z.B. vorhandene oder fehlende Kindgemäßheit der Instruktion, der Aufgabenstellung usw.) bei der Wahl des Testinstrumentes berücksichtigen müssen. Aspekte der Praktikabilität sind keinesfalls unbedeutend, sie stellen jedoch keine „essentiellen“ Merkmale (wie die Testgütekriterien) dar, zudem eignet ihnen gewöhnlich eine größere Evidenz, so daß wir uns auf eine ausführlichere Erörterung der drei Hauptkennfunktionen pädagogisch-psychologischen Testens beschränken können.

#### *Objektivität*

Objektivität eines Tests oder eines beliebigen Standardmeßverfahrens liegt dann vor, wenn die auf Seite 57ff. skizzierten (Standard-)Untersuchungsbedingungen erfüllt sind. Im einzelnen gehören dazu präzise Hinweise für die Gestaltung der Testsituation i.w.S., für die Testleiter und Probanden eindeutig formulierte Instruktionstexte mit genauen Protokollanweisungen, verbindliche Richtlinien für die Testauswertung und Interpretation der Ergebnisse, so daß interpersonelle Übereinstimmung der Testleiter, d.h. Unabhängigkeit der Datenauswertung und Befunddeutung von der Person des Interpreten gewährleistet ist. Um die Stellung des Individuums (Probanden) – z.B. hinsichtlich seiner intellektuellen Leistungen – innerhalb seiner Bezugsgruppe (Alters-, Geschlechts-, Schul- oder Klassengruppe) objektiv bestimmen zu können, bedarf es weiterhin entsprechender – auf die jeweilige Gruppe bezogener – Leistungsmaßstäbe (Testnormen), die im Testmanual in Form von Tabellen oder Kurvendarstellungen beigelegt sind. Das Normenproblem wird uns noch eingehend beschäftigen (s. S. 75ff.). Die Objektivität eines Tests bezieht sich also gleichermaßen auf die Testdurchführung wie auf die *Auswertung* und *Interpretation* von Testbefunden: Der Grad intersubjektiver Übereinstimmung in diesen Punkten gibt Auskunft über die jeweilige Objektivität eines Tests. Im weiteren Sinne ist hier die gesamte psychosoziale Testkonstellation einschlägig (vgl. S. 59ff.).

#### *Reliabilität und Standardmeßfehler*

Tests dienen einem bestimmten Untersuchungszweck. Intelligenztests sollen beispielsweise Auskunft über intellektuelle Fähigkeiten, diverse Begabungen oder Schuleignungen geben. Dabei stellt sich die Frage nach der Treffsicherheit der gewünschten bzw. erhaltenen Informationen, d.h. *ob* und *wie genau* Tests tatsächlich messen (können); der erste Aspekt betrifft die Validität einer Messung, der zweite die Reliabilität. Bei der *Reliabilität eines Tests* ist gefragt, inwieweit einmaliges oder punktuell

erfaßtes Testverhalten (also eine Verhaltens*stichprobe*) in bezug auf Verhaltensstichproben des gleichen Tests bei anderer Gelegenheit verallgemeinert werden darf. Oder man könnte fragen, *wie zuverlässig* das einmalige Testleistungsverhalten das durchschnittliche (nicht via Test erfaßte) Leistungsverhalten des Probanden repräsentiert bzw. vorhersagt. M.a.W.:

Die Aussage eines (Intelligenz-)Tests muß für das Individuum hinsichtlich bestimmter (z.B. intellektueller) Fähigkeiten oder Merkmalsausprägungen *kennzeichnend* sein. Die Reliabilität betrifft also die *formale Genauigkeit* von Testwerten bzw. eines beliebigen Meßinstrumentes. Eine häufig verwendete Definition lautet deshalb kurz und bündig: *Reliabilität ist die Genauigkeit, mit der ein Test das mißt, was er tatsächlich mißt – ohne Rücksicht auf den jeweiligen Meßinhalt* (vgl. Validität).

Die im Fachjargon häufig benutzte Formel „Reliabilität eines Tests“ könnte leicht dazu verführen, die Reliabilität als allgemeine Eigenschaft eines Tests zu verstehen. Streng genommen bezeichnet jedoch *Reliabilität* nicht eine Testeigenschaft, sondern eine *Eigenschaft einer Reihe von Messungen*, die mit dem Test an einer bestimmten Population vorgenommen wurden (Michel 1964, S. 35). Eingeführt wurde der Begriff von Spearman (1904), Brown (1910) und Stern (1911). Wilde und Lienert sprechen von *Zuverlässigkeit*, Hofstätter von *Verlässlichkeit*, Meili von *Stabilität*.

Die angeführten Begriffe sind mehr als bloße Synonyme, sie symbolisieren teilweise unterschiedliche Aspekte der Meßgenauigkeit. „Reliabilität“ (engl. reliability) fungiert hier als Oberbegriff für eine Reihe von Konzepten, unter denen *Termini* wie (*innere*) *Konsistenz*, (*Zeit*-)*Stabilität* oder (*Standard*-)*Meßfehler* die wichtigsten Kriterien der Reliabilität ausmachen. Dazu bedarf es zusätzlicher Erläuterungen.

Alle Reliabilitätsarten basieren auf zwei (hypothetischen) Voraussetzungen: a) der Annahme relativ überdauernder Merkmalskonstanz (siehe dazu die früheren Ausführungen auf Seite 31ff.), b) dem Postulat der Testobjektivität bzw. Standardisierung der Testsituation i.w.S. (s. S. 63 u. 57ff.). Sofern diese Bedingungen erfüllt sind, gehen beobachtete Leistungsschwankungen im Test zu Lasten des Testinstrumentes, sind also Ausdruck mehr oder weniger mangelnder Zuverlässigkeit. In jedem Testresultat stecken somit „wahre“ und „Fehler“-Varianzanteile.

Wir betonten bereits den Tatbestand „interindividueller Differenzen“, d.h. die Tatsache, daß menschliche Eigenschaften und Fähigkeiten (inter)individuell verschieden ausgeprägt sind. Der darauf beruhende Varianzanteil wird in der Teststatistik „wahre“ Varianz ( $s_{\infty}^2$ ) genannt. Im Intelligenztest etwa meint die Bezeichnung  $s_{\infty}^2$  die tatsächlich vorhandene (nicht gemessene) „Intelligenz“. „Wäre es möglich, das Persönlichkeitsmerkmal unabhängig von allen inneren und äußeren Bedingungen zu messen, so würde sich für eine definierte Stichprobe stets dieselbe „wahre“ Varianz ergeben. Aufgrund allgemeiner experimenteller Erfahrungen weiß man jedoch, daß es bei jeder – auch der physikalischen – Messung Fehlerquellen gibt, zu deren Lasten ein Teil der beobachteten Gesamtvarianz geht. Dieser Teil sei „Fehler“-Varianz genannt und mit  $s_e^2$  bezeichnet“ (Lienert 1967, S. 211). Sofern es sich stets um dieselbe Stichprobe handelt, gilt nun folgende Beziehung:

$$s_x^2 = s_{\infty}^2 + s_e^2$$

Mit anderen Worten: Die im Test gemessene bzw. *beobachtete Varianz* ( $s_x^2$ ) der ermittelten Testrohwerte setzt sich aus „wahrer“ Varianz ( $s_{\infty}^2$ ) und „Fehler“-Varianz ( $s_e^2$ ) zusammen. Der *Reliabilitätskoeffizient* ( $r_{tt}$ ) als Kennwert der Meßgenauigkeit ist nun allgemein durch das Verhältnis der wahren Varianz zur Gesamtvarianz (beobachtete Varianz) bestimmt:

$$r_{tt} = \frac{s_{\infty}^2}{s_x^2} \frac{\left[ \begin{array}{l} \text{„wahre“, d.h. nicht zufallsbedingte interindividuelle Unterschiede} \\ \text{der Merkmalsausprägung (z.B. der Intelligenz)} \end{array} \right]}{\left[ \text{Gesamtstreuung} \right]}$$

Da  $s_{\infty}^2 = s_x^2 - s_e^2$  ist, läßt sich obige Gleichung folgendermaßen umformen:

$$r_{tt} = 1 - \frac{s_e^2}{s_x^2} \quad (\text{nach Lienert, loc. cit.})$$

Hieraus wird ersichtlich, daß die Reliabilität mit zunehmender Fehlervarianz abnimmt. Die Fehlervarianz kann zu Lasten „innerer“ Fehler (*mangelnder Konsistenz*) des Testinstruments, z.B. mangelnder Homogenität der Testaufgaben, mangelnder Auswertungsobjektivität) oder/und zu Lasten „äußerer“ Störvariablen (Variation der Testdurchführungsbedingungen, z.B. Veränderung der Testinstruktion, des Testleiterverhaltens, der Motivation seitens des Probanden u.ä. Situationsvariablen) gehen. Entsprechend müssen die *Methoden zur Reliabilitätsbestimmung* gewählt werden. Am gebräuchlichsten sind folgende vier Verfahren:

1. *Die Retest-Methode.* Gemeint ist hier die Durchführung einer zweimaligen Untersuchung mit demselben Test an derselben Population. Anschließend werden die Testergebnisse aus Erst- und (in einem zeitlichen Intervall erfolgter) Zweitestung korreliert, um den Zusammenhang beider Meßvariablenreihen zu ermitteln. Der Korrelationskoeffizient kann wieder Werte zwischen  $-1$  und  $+1$  (100prozentige Deckungsvarianz) annehmen. Je höher der Koeffizient ausfällt, desto besser ist die *Stabilität* der Testergebnisse über einen gewissen Zeitraum hinweg und damit die Wiederholungs- oder Retest-Reliabilität. Dieses Verfahren darf aber nur angewandt werden, wenn keine Wiederholungseinflüsse (z.B. Lerneffekte) auftreten bzw. solche Einflüsse sich bei allen Probanden gleichmäßig auswirken.
2. *Die Paralleltest-Methode.* Diese Technik ist nur bei Tests mit gleichwertigen Parallelformen anwendbar. Hierbei wird dieselbe Pbn-Gruppe entweder in unmittelbarer oder zeitlich getrennter Reihenfolge untersucht und anschließend wieder die Korrelation bestimmt. Häufig wird das als die „beste“ Reliabilitätsbestimmung betrachtet. Ein Teil der Einflüsse wie beim Test-Retest dürfte jedoch auch hier wirksam sein.
3. *Die Halbierungsmethode (Split-Half-Reliabilität).* Diese Methode findet wohl am meisten Verwendung. Ihre Anwendung ist nur dann erlaubt, wenn der Test in faktoriell gleichwertige (äquivalente) Hälften geteilt werden kann und verbietet sich im allgemeinen bei Speed-Tests<sup>40</sup>.
4. *Die Analyse der Inter-Item-Konsistenz.* Diese Methode stellt eine Verallgemeinerung des Halbierungsverfahrens dar, wobei die Aufteilung in so viele Teile möglich ist, wie Items (Testaufgaben) vorliegen. Im übrigen gelten ähnliche Voraussetzungen wie bei der Split-Half-Methode.

Die *Retest-Reliabilität* bestimmt den Grad der Genauigkeit, mit der Testwerte überdauernde psychische Unterschiede der Individuen aussagen. Nun werden realiter immer momentane Einflüsse der „äußeren“ Testsituation in das Testergebnis mithereinspielen. Die daraus resultierende Abweichung der „wahren“ Testwerte heißt Fehlervarianz. Diese wird durch Stabilitätsuntersuchungen nach der Retestmethode geschätzt (siehe oben).

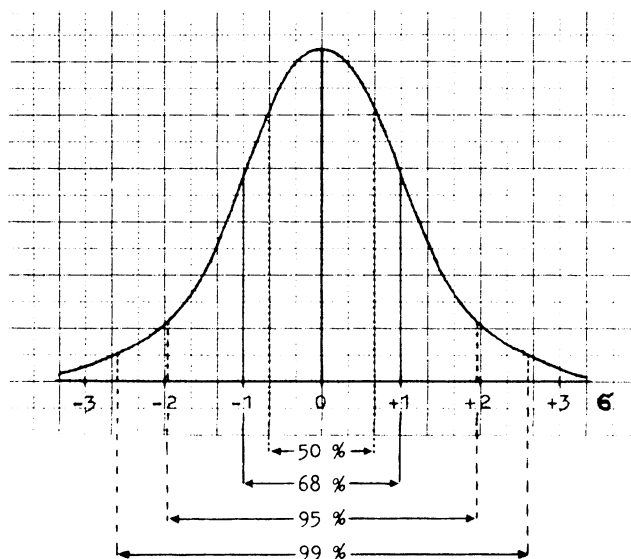
Eine andere Ursache irreliabler Intelligenz-/Leistungsmessung liegt in der *begrenzten Anzahl* von Testaufgaben und/oder in deren (eingeeengten) Auswahlkriterien. Zum ersten Aspekt wäre zu sagen, daß eine größere Aufgabenzahl im allgemeinen die Zuverlässigkeit von Tests steigert. Im Hinblick auf die *Itemauswahl* gilt die Forderung nach maximaler Repräsentanz in bezug auf das Universum möglicher Testitems. M.a.W.: Jeder Test enthält – schon aus zeitlich-ökonomischen Gründen – nur eine begrenzte Zahl von (generell akzep-

<sup>40</sup>In der Testliteratur unterscheidet man zwei Grundtypen: *Speed- oder Geschwindigkeits-Tests* (z.B. der Pauli-Test) und *Power- oder Niveau-Tests* (z.B. der Progressive Matrizen-Test von J.C. Raven). Häufig finden auch Mischtypen Verwendung (vgl. z.B. HAWIK bzw. HAWIE nach D. Wechsler, LPS bzw. PSB von W. Horn, Frankfurter Wortschatztest WST von H. Anger et al.).

tierbaren) Aufgabenmöglichkeiten. Bei der Auswahl ist nun darauf zu achten, daß die Testaufgaben keinen potentiellen Probanden bevorzugen versus benachteiligen, soweit inhaltlich-strukturelle Aufgabenmomente betroffen sind. Gerade bei *verbalen* Testaufgaben stößt die Verwirklichung dieser Forderung häufig auf erhebliche Schwierigkeiten (siehe unsere früheren Erörterungen auf Seite 38ff.). Irreliabilitätsgrade zu Lasten der Testaufbaustruktur werden durch die Split-Half-Technik sowie Interitem-Konsistenz-Analyse erfaßt.

In der amerikanischen Literatur taucht zuweilen die Unterscheidung zwischen *Äquivalenz*-Koeffizienten (als Maß für die Halbierungs-Reliabilität) und *Stabilitäts*-Koeffizienten (als Maß für die Retest-Reliabilität) auf. Im Grunde sind beide Aspekte der Testzuverlässigkeit wichtig und sollten – soweit nur möglich – jeweils kontrolliert werden. Die rigoroseste Reliabilitätsprüfung stellt indessen eine *Kombination von Paralleltest- und Retest-Technik* dar. Hierbei wird in der Wiederholungstestung (etwa nach 2 Wochen) auf die in der Ersttestung (jeweils bei den betr. Probanden) nicht applizierte Paralleltestform zurückgegriffen. Auf diese Weise bekommt man Faktoren der Äquivalenz und der Stabilität gleichermaßen in den Griff. Je mehr Aspekte der Reliabilität erfaßt werden, um so besser wird man der Alltagsrealität gerecht, unter deren Bedingungen die meisten Testuntersuchungen später ablaufen. Ausführlicher informiert Michel (loc. cit.) über die zuletzt angesprochenen Probleme der Reliabilitätskontrolle.

Neben den Reliabilitätskoeffizienten gibt der *Standardmeßfehler* (oder kürzer: *Meßfehler*) über die Zuverlässigkeit eines Meßinstrumentes, z.B. eines Tests, Auskunft. Die Funktion des Meßfehlers mag uns ein Exkurs in die Stichprobenstatistik verdeutlichen. In der empirischen bzw. experimentellen Forschung ist man normalerweise auf Stichprobenerhebungen angewiesen. Unter einer *Stichprobe* (Sample) versteht man eine relativ kleine Gruppe von Angehörigen einer wohldefinierten (größeren) Gruppe oder Population, die



Grafik 1. Standardabweichungen ( $\sigma$ ) als Parameter der normierten Gaußschen Verteilung sowie Akzeptationsbereiche (Wahrscheinlichkeiten in %) versus Irrtums-(Rest-)Wahrscheinlichkeiten für verschiedene Signifikanzniveaus – jeweils als Flächenanteile unter der Normalkurve.

für bestimmte Untersuchungszwecke nach vorher festgelegten Kriterien, z.B. des Zufalls oder der Repräsentanz (relevanter Merkmale), ausgewählt wird. *Population* oder *Kollektiv* oder *Grundgesamtheit* ist demgegenüber definiert als Gesamtheit aller Merkmalsträger, d.h. aller Individuen, die gleiche Merkmale oder Merkmalskombinationen tragen. Eine Population ist also keine natürliche Gruppe; sie entsteht vielmehr aufgrund einer Definition. So können beispielsweise alle Viertklässler der BRD oder alle Oberstufenschüler westdeutscher Blindenschulen oder alle männlichen Stotterer im Schulpflichtalter etc. als (Untersuchungs-)Populationen bestimmt werden. Statistische Kennwerte (z.B. Mittelwerte, Varianzen, Prozentquoten, IQ-Werte u.ä.), die sich auf die Population beziehen, nennt man *Parameter*, solche, die sich auf Stichproben beziehen, *Statistiken*.

Jede Statistik ist mit Fehlern, die sich aus dem Stichprobencharakter einer Messung (jeder Messung) erklären lassen, behaftet. So ist etwa der an Hand einer Zufallsstichprobe gefundene IQ-Mittelwert bei lernbehinderten Kindern im Grundschulalter *nicht* identisch mit dem „wahren“ Mittelwert der Grundgesamtheit „Lernbehinderte Unterstufenschüler“. Vielmehr steht zu erwarten – und entsprechende empirische Untersuchungen bestätigen dies –, daß die Statistik (hier: IQ-Durchschnitt der Stichprobe) mehr oder weniger stark vom Parameter (IQ-Durchschnitt der Population) abweicht. Bei mehreren Stichprobenerhebungen verteilen sich nun die Stichproben-Mittelwerte derart um den „wahren“ Mittelwert, daß die bekannte Normal- oder Gaußsche Kurve (vgl. Grafik 1) entsteht. Unter der Voraussetzung genügend häufiger Schüler-Stichproben wäre der IQ-Mittelwert der *Stichprobenverteilung* <sup>41</sup> gleich dem IQ-Mittelwert des Kollektivs und läge genau in der Mitte der Gaußschen Kurve; wäre er größer als der „wahre“ Mittelwert, so läge er rechts von der Mitte, wäre er kleiner, so läge er links von der Mitte der Normalkurve. In unserer Grafik sind die Abweichungen vom Mittelwert in Einheiten der sog. *Standardabweichung*  $\sigma$  bzw.  $s = \sqrt{\sum (M - X)^2 / N}$  ausgedrückt; siehe noch Hofstätter & Wendt 1966, S. 38 f.

In der Forschungspraxis, wo man zumeist auf eine einzige Stichprobe angewiesen ist, interessiert vor allem die Frage, ob bzw. wie man Parameter schätzen kann, d.h. die Möglichkeit, vom Mittelwert der Stichprobe – um bei unserem Anwendungsbeispiel zu bleiben – auf den wahren Mittelwert zu schließen. Tatsächlich kann man mit Hilfe des sog. *Standardfehlers*, d.h. der Standardabweichung der Stichprobenverteilung ( $\sigma_M$ ) die gewünschte Parameterschätzung vornehmen <sup>42</sup>. Der *Standardfehler* (Standard Error of the Mean) bestimmt sich nach der Formel  $\sigma_M = SE = s / \sqrt{N - 1}$ , wobei  $\sigma_M$  bzw. SE als Symbol für den Standardfehler stehen,  $s$  die Standardabweichung der Stichprobe symbolisiert und  $N$  die Größe der Stichprobe angibt. Über den so berechneten Standardfehler des arithmetischen Mittels lassen sich dann die *Konfidenzintervalle* (CL = Confidential Limits) jeder beliebigen Statistik abgrenzen. Damit ist jener Bereich, in dem der „wahre“ Wert liegen muß, gemeint. Speziell für die Berechnung der Vertrauensgrenzen des Mittelwertes der Population ( $M_{pop}$ ) gilt demnach die Beziehung:  $M_{pop} = M \pm z \cdot \sigma_M$ , wobei sich  $M$  auf die Stichprobe bezieht,  $z = X - M / \sigma$  und  $\sigma_M = s / \sqrt{N - 1}$  ist.

Analog zur Bestimmung des Vertrauensbereiches einer Statistik läßt sich über den (*Standard-*)Meßfehler ( $s_e$ ) das Vertrauensintervall eines Testwertes definieren. Die Axiomatik der klassischen Meßfehlertheorie basiert auf folgenden Voraussetzungen (nach Pawlik 1970):

1. *Existenzaxiom*: Zu jedem beobachteten (gemessenen) Wert  $X$  existiert ein „wahrer“ Wert, d.i. psychologisch eine bestimmte individuelle Merkmalsausprägung. Diese wird

<sup>41</sup>Es ist zu beachten, daß es sich hier nicht um individuelle Werte, sondern um (Stichproben-) Mittelwerte in der Verteilung handelt.

<sup>42</sup>Man spricht in diesem Zusammenhang von „Fehler“ (und nicht von „Abweichung“), weil sich  $\sigma_M$  auf die M-Differenzen zwischen Stichprobe und Population (und nicht auf individuelle Abweichungen vom Gruppenmittelwert bzw. interindividuelle Differenzen) bezieht.

als Konstante – wenigstens über einen gewissen Zeitraum hinweg – angenommen (vgl. dazu unsere früheren Ausführungen auf Seite 31ff.).

2. *Fehleraxiom:* Der Meßfehler einer Messung ist eine Zufallsvariable. Für diese gilt, daß die Summe bzw. das arithmetische Mittel (M) der Fehlerwerte den Wert Null ergibt.
3. *Verknüpfung axiom:* Der beobachtete (gemessene) Wert setzt sich additiv aus „wahrem“ Wert und Fehlerwert zusammen (siehe auch Seite 64f.). Darüberhinaus wird meistens angenommen, daß die Meßfehler *normal* verteilt seien (siehe oben).

Ein kleines Gedankenexperiment möge auch hier wieder die Zusammenhänge veranschaulichen. Nehmen wir an, wir würden einen Grundschüler der 4. Klasse innerhalb des letzten Quartals mehrfach bzw. beliebig oft mit dem gleichen Intelligenztest untersuchen, wobei die Frage der Schuleignung für die Sekundarstufe I im Vordergrund des Interesses stehen soll. Inwieweit können wir nun – von der Frage der Validität des Tests (siehe unten) einmal abgesehen – dem Testergebnis vertrauen? Nach den vorausgegangenen Erörterungen zum Problem des Standardfehlers des arithmetischen Mittels resp. des Stichprobenfehlers erwarten wir wiederum, daß die einzelnen Testwerte – in der Form einer Normalverteilungskurve – um den „wahren“ Testwert des Probanden streuen, also mehr oder weniger stark (positiv versus negativ) vom „tatsächlichen“ IQ-Wert des Probanden abweichen. Diese Abweichungen oder *Meßfehler* (vgl. Fußnote 42) sind das Resultat mannigfacher (äußerer und innerer) Störeinflüsse, wie sie etwa in jeder konkreten Testsituation stecken. Die hic et nunc erhaltenen Testergebnisse repräsentieren (noch) nicht die zu erwartende intellektuelle Schultüchtigkeit unseres Probanden, sie müssen vielmehr als Summe der wahren Varianz und der Fehlervarianz betrachtet werden. Aus den oben dargestellten Axiomen kann jedoch abgeleitet werden, daß das *arithmetische Mittel* der beobachteten Testwerte gleich ist dem arithmetischen Mittel der wahren Testwerte (in unserem Beispiel der „wahren Schuleignung“), womit der Weg zur Isolierung der Fehlervarianz eröffnet wäre. Da im Regelfalle aufgrund ein- oder zweimaliger (selten mehrmaliger) Testung intelligenzdiagnostische bzw. schuleignungsprognostische Urteile gefällt werden müssen, ist man auch hier wiederum auf die Bestimmung der Konfidenzintervalle, also jener Grenzen, innerhalb derer der „wahre“ Wert liegt, angewiesen. Dabei kommt einem die bekannte Standardabweichung zugute, welche die *Wahrscheinlichkeit* (unter der Normalkurve) anzeigt, mit der ein „wahrer“ Testwert innerhalb bestimmter Grenzen liegen wird. Aus Grafik 1 (S. 66) ist zu sehen, daß dieser mit 68% Wahrscheinlichkeit innerhalb des  $\pm 1$ -Sigma-Bereichs liegen muß, mit 95% Wahrscheinlichkeit innerhalb von  $\pm 1,96 \sigma$  und mit 99% Wahrscheinlichkeit innerhalb  $\pm 2,58 \sigma$ , jeweils vom Mittelwert (Sigma = 0) aus gerechnet. In der älteren Statistik wird manchmal noch der „*wahrscheinliche*“ Fehler (WF) angeführt. Dieser bezieht sich auf jene Bereichsgrenzen, innerhalb und außerhalb derer *gleiche* Auftretenshäufigkeit (nämlich je 50% Wahrscheinlichkeit) besteht, also  $\pm 0,6745 \sigma$ . Für die Beziehung von Standard(meß)fehler und wahrscheinlichem Fehler (Probable Error) gilt deshalb:  $WF = PE = 0,6745 \cdot \sigma_M$ .

Je nach aktuellem Sicherheitsbedürfnis, d.h. gewünschter Wahrscheinlichkeit bzw. testdiagnostischer Verlässlichkeit versus Fehlerrisikobereitschaft, setzt man die entsprechenden Sigma-Werte zum Reliabilitätskoeffizienten des benutzten Tests in Beziehung, indem man den *Standardmeßfehler* ( $s_e$ ) jetzt miteinbezieht. Dieser wird in Rohpunkt-Einheiten für den betr. Test nach der Formel  $s_e = s_x \sqrt{1 - r_{tt}}$  errechnet resp. vom Autor im Test-Manual mitgeteilt. Über den Standardmeßfehler lassen sich so – unter Beachtung des jeweiligen Signifikanzniveaus bzw. der Irrtumswahrscheinlichkeit (probabilitas) – folgende *Vertrauensbereiche* oder *Konfidenzintervalle* für beliebige Testwerte ermitteln:

1.  $CL_X = X \pm 1 \cdot s_e$  (für  $p = 32\% - 32\%$ -Niveau)
2.  $CL_X = X \pm 1,96 \cdot s_e$  (für  $p = 5\% - 5\%$ -Niveau)
3.  $CL_X = X \pm 2,58 \cdot s_e$  (für  $p = 1\% - 1\%$ -Niveau)

Zwei Rechenbeispiele mögen das Vorgehen konkretisieren. Die Anwendung des Progressiven Matrizen-Tests (PMT) bei tauben Kindern der 1. Grundschulklasse erbrachte eine Standardabweichung von  $s_x \sim 4$  (vgl. Tab. Ib im Anhang dieses Buches) und eine Retest-Reliabilität von  $r_{tt} = 0.98$  (Heller 1967 b, S. 227). Wie vertrauenswürdig ist nun ein beliebiger Rohwert des PMT in der Anwendung bei gehörlosen Erstkläßschülern? Zunächst errechnen wir nach oben angegebener Formel den Standardmeßfehler:  $s_e = 4 \sqrt{1 - 0.98} = 4 \cdot 0.14 \sim 1$ . Nach der CL $_X$ -Formel für 5prozentige Irrtumswahrscheinlichkeit ergibt sich demnach:  $CL_X = X \pm 1.96 \cdot 1 = X \pm \sim 2$ . M.a.W.: Die wahre Intelligenztestleistung eines tauben Erstkläßlers liegt – bei einer Irrtumswahrscheinlichkeit von maximal 5% – höchstens 2 Rohpunkte (RP) *über* oder *unter* dem jeweils gemessenen PMT-Wert (RP). Sofern eine Verringerung der Irrtumswahrscheinlichkeit dieser Aussage wünschenswert erscheint, kann man auch obige CL $_X$ -Formel Nr. 3 verwenden. In diesem Fall würde sich allerdings das Konfidenzintervall von 4 RP auf 5 RP vergrößern.

Auf der Grundlage von Testnormen (Standardwert-Normen), wenn also die Streuungsmaße vereinheitlicht sind, lassen sich auch die Standardmeßfehler verschiedener Tests miteinander vergleichen. Dies ist für die diagnostische Praxis (i.e.S. die Interpretation von Testbefunden) von Bedeutung. Außerdem kann man dann – die Kenntnis der betr. Testreliabilität vorausgesetzt – den Meßfehler für häufiger benutzte Tests leicht im Kopf ausrechnen. So beträgt für IQ-Normen (Wechsler- oder Abweichungen-IQ) die Standardabweichung  $s_{IQ} = 15$ , für Centil-Normen die Standardabweichung  $s_C = 2$  oder für T-Normen die Standardabweichung  $s_T = 10$ . Lienert (1967, S. 454) bezeichnet den *Meßfehler für Standardwerte* mit dem Symbol  $s'_e$ , um den Unterschied zum Meßfehler  $s_e$ , der sich auf Rohpunkt-Einheiten bezieht, anzuzeigen. Für IQ-Normen ist demnach  $s'_e = 15 \sqrt{1 - r_{tt}}$  (z.B. HAWIK, PMT tb./resthör./schwerhör.), für C-Normen  $s'_e = 2 \sqrt{1 - r_{tt}}$  (z.B. LPS, PSB), für T-Normen  $s'_e = 10 \sqrt{1 - r_{tt}}$  (z.B. FTU 4-6). Unser zweites Rechenbeispiel bezieht sich nun auf den Meßfehler  $s'_e$ .

Eine Schülerin der 7. Hauptschulklasse (Alter 13;2) sei mit dem LPS (Leistungsprüfsystem von Horn) getestet worden und habe einen Gesamtleistungswert von 5,6 C erhalten. Für die Centil-Normen des LPS ermittelten wir bei einer durchschnittlichen (minderungskorrigierten) Parallel-Retest-Reliabilität von 0.943 für 13jährige Schüler (vgl. Tent 1969, S. 89) einen Meßfehler von  $s'_e \sim 0,3$  C. Das Konfidenzintervall für den Testwert 5,6 C beträgt somit  $CL = 5,6 \pm 1,96 \cdot 0,3 = 5,6 \text{ C} \pm \sim 0,6 \text{ C}$  (für  $p = 5\%$ ) beziehungsweise  $CL = 5,6 \pm 2,58 \cdot 0,3 = 5,6 \text{ C} \pm \sim 0,8 \text{ C}$  (für  $p = 1\%$ ). Mit anderen Worten: Der wahre LPS-Wert unserer Probandin liegt zwischen 5,0 C und 6,2 C beziehungsweise zwischen 4,8 C und 6,4 C, was nach der Transformationstabelle von Testnormen (vgl. Tab. VIII im Anhang) dem IQ-Bereich von 100 bis 109 beziehungsweise 99 bis 110 entspricht.

Sofern keine Standardwert-Normen, sondern (nur) Prozentränge (PR) als Normen zur Verfügung stehen, was bei Schultests häufiger der Fall ist – siehe auch Tab. IIa, IV, VIc und VIId im Anhang dieses Werkes –, muß man den *Standard(meß)fehler von Prozentzahlen oder Prozenträngen* (PR) verwenden. Die Formel hierfür lautet:  $\sigma\% = \sqrt{\frac{P \cdot Q}{N}}$ ,

wobei die Symbole  $\sigma\%$  = Standard(meß)fehler der %-Zahl oder PR, P = empirisch ermittelte Prozentzahl (PR), Q = 100 – P und N = Stichprobengröße (Größe der Eichstichprobe) = Prozentuierungsbasis bedeuten. Entsprechend werden über den Standard(meß)fehler von %-Zahlen bzw. PR wiederum die Vertrauensgrenzen einer Statistik oder eines PR-Wertes im Test bestimmt nach der Formel:

(4)  $CL\% = P \pm z \cdot \sigma\%$ , wobei für die Wahl des z-Wertes (gewöhnlich 1,00 oder 1,96, seltener 2,58) das gewünschte Sicherheitsniveau (68 : 100 oder 95 : 100 bzw. 99 : 100) maßgebend ist.

Zuweilen genügt es auch, die Vertrauensbereiche (lediglich) für das sog. *mittlere Quartil* (Q) und/oder den *Median* (Med) zu bestimmen. Für die Berechnung des *Standard(meß)fehlers des mittleren Quartils* ( $\sigma_Q$ ) gibt Fröhlich (1965, S. 40) zwei Möglichkeiten an:



$$a) \sigma_Q = 0,786 \cdot \sigma'_M \quad b) \quad Q = \frac{1,17 \cdot Q}{\sqrt{N}}, \text{ wobei } Q = \frac{Q_3 - Q_1}{2} \text{ ist } 43.$$

Die Differenz zwischen  $Q_3$  und  $Q_1$  dividiert durch 2 ergibt also hier den Wert  $Q$ , d.h. das mittlere Quartil;  $N$  bezieht sich wieder auf die Stichprobengröße (Prozentuierungsbasis) und 1,17 bzw. 0,786 sind Konstanten.

Der *Standardfehler des Median* ( $\sigma_{Med}$ ) kann nach folgenden Formeln geschätzt werden:

$$a) \sigma_{Med} = 1,253 \cdot \sigma'_M \quad b) \sigma_{Med} = \frac{1,253 \cdot \sigma}{\sqrt{N}} \quad (\text{nach Fröhlich})$$

$$c) \sigma_{Med} = \frac{1,858 \cdot Q}{\sqrt{N}} \quad (\text{nach Süllwold})$$

Über die so geschätzten Standardfehler ( $\sigma_Q$  bzw.  $\sigma_{Med}$ ) können dann nach der CL%-Formel Nr. 4 die Konfidenzintervalle für das mittlere Quartil und den Median errechnet werden. Dies empfiehlt sich beispielsweise, wenn die Normentabellen im Appendix Nr. IIb, V, VIa, VIb und VII oder ähnliche benutzt werden. Allerdings sollten  $\sigma_Q$  und  $\sigma_{Med}$  nur dann berechnet werden, wenn die statistischen Voraussetzungen (Intervallskala, Normalverteilung) für die Verwendung von  $\sigma_M$  bzw.  $s_e$  oder  $s'_e$  nicht gegeben sind und/oder  $N$  kleiner als 100 ist resp. die empirisch ermittelten %-Zahlen nicht außerhalb von 10 und 90 liegen, da sonst u.U. größere Schätzfehler in die Berechnung eingehen.

Unsere Ausführungen über den Standard(meß)fehler resümierend halten wir somit fest, daß über  $\sigma_M$  bzw.  $s_e$  und  $s'_e$  oder  $\sigma_Q$  bzw.  $\sigma_{Med}$  und  $\sigma_Q$  jeweils das Konfidenzintervall eines „wahren“ Testwertes bestimmt und damit allgemein Aussagen über die Zuverlässigkeit von Messungen ermöglicht werden. Die Ermittlung von Meßfehler und Vertrauensbereich (z.B. eines Intelligenztestwertes) dient aber nicht nur der genaueren Bestimmung je individuell ausgeprägter Fähigkeits- und Merkmalsgrade, mit ihrer Hilfe können darüber hinaus inter- und intraindividuelle Testwertunterschiede und damit Intelligenz- oder Begabungsunterschiede auf Zufallsbedingtheit versus Überzufälligkeit (Signifikanz) hin kontrolliert werden. Zu diesem Zweck wird in manchen Schultests das Vertrauensintervall jedes Testscores sogleich als Normen-Band in die Normentabelle eingearbeitet, wobei man sich in der Regel auf die Vertrauensgrenzen für  $p = 32\%$  oder – bei einseitiger Fragestellung –  $16\%$  festlegt <sup>44</sup>. Seltener wird heute noch der „wahrscheinliche“ Fehler zur Abgrenzung des Konfidenzintervalls (vgl. S. 68) verwandt, etwa in der HAWIK-Profilauwertung zur Ermittlung lernbehinderter Kinder nach Reinartz & Seifart (1970).

<sup>43</sup> Das 3. Quartil ( $Q_3$ ) fällt mit dem 75. PR und das 1. Quartil ( $Q_1$ ) mit dem 25. PR zusammen. Entsprechend ist das 2. Quartil ( $Q_2$ ) oder der Zentralwert oder Median mit dem 50. PR identisch.

<sup>44</sup> Zweiseitige Fragestellung berücksichtigt die „Rest“-Wahrscheinlichkeiten zu beiden Seiten (links und rechts) der Verteilungskurve, also hier 2mal  $16\% = 32\%$  (siehe noch S. 77, Graf. 2). Mitunter genügt jedoch die *einseitige* Fragestellung, etwa bei der Ermittlung von Lernbehinderung i.S. der Sonderschulbedürftigkeit oder bei der Erfassung der Gymnasialeignung mit Hilfe von IQ-Tests, wobei dann nur *eine* Restwahrscheinlichkeitshälfte (bei der Erfassung der Lernbehinderung nur die „obere“ versus bei der Ermittlung der Gymnasialeignung nur die „untere“ Bereichsgrenze) kritisch wird. M.a.W.: Es sind drei Möglichkeiten von  $p$  zu unterscheiden: die Wahrscheinlichkeit der positiven und negativen Abweichung (2seitiger Signifikanztest) und die Wahrscheinlichkeit der positiven versus negativen Abweichung (1seitiger Signifikanztest). Bei der *einseitigen* Fragestellung ist also nur die Wahrscheinlichkeit entweder der positiven oder der negativen Abweichung vom „wahren“ Testwert des Probanden interessant, was unter Verwendung des o.a. Normenbandes eine Irrtumswahrscheinlichkeit von jeweils  $p = 16\%$  bedeutet.

Nehmen wir beispielsweise an, drei Viertklässchüler hätten folgende Ergebnisse im FTU 4–6 (Fremdsprachen-Eignungstest für die Unterstufe, d.h. 4. bis 6. Klassen, nach Correll & Ingenkamp) erzielt:

Schüler(in)	RP	PR-Band	T-Wert-Band	M <sub>T</sub>	Signifikanz-Niveau
Bernd	67	76–88	57–62	60	p = 32 %
Renate	73	86–96	61–67	64	p = 32 %
Dieter	58	58–73	52–56	54	p = 32 %

Die Berücksichtigung des Konfidenzintervalls erleichtert zum einen die genauere (d.h. mit einer best. Zuverlässigkeitswahrscheinlichkeit, im obigen Beispiel mit 68 zu 100 sichere) Lokalisation des wahren Fremdsprachentestwertes (der tatsächlichen Fremdspracheneignung) für den einzelnen Probanden, zum andern aber wird dadurch die Sicherheit des diagnostischen Urteils über evtl. *interindividuelle Differenzen* erhöht. So können wir jetzt – mit einer möglichen Irrtumswahrscheinlichkeit von maximal 32 % – die Frage entscheiden, *ob sich die FTU-Ergebnisse unserer drei Schüler nur zufällig oder tatsächlich (signifikant) unterscheiden*. Differente Testergebnisse zweier Individuen unterscheiden sich nur zufällig (nicht signifikant), wenn sich die Normenbänder, in denen die gefundenen Individualscores liegen, überlappen. Liegen keine Überschneidungen vor, kann man die beobachteten (gemessenen) Testleistungsdifferenzen als *tatsächliche* (signifikante) interindividuelle Intelligenz-, Leistungs- oder Merkmalsunterschiede interpretieren. In unserem Fallbeispiel unterscheiden sich demnach Bernd und Renate in Bezug auf ihre Fremdsprachengabung – nach dem FTU 4–6 – nur zufällig, während beide dem Pb Dieter signifikant überlegen sind. Analog zu diesem Vorgehen, interindividuelle Differenzen zu interpretieren, kann man auch *intraindividuelle* Testleistungsunterschiede mit Hilfe des Konfidenzintervalls diagnostizieren. So lassen sich etwa auf diese Weise bedeutsame Fortschritte (versus Rückschritte) in der individuellen Begabungs*entwicklung* eines Schülers feststellen. Weitere Beurteilungsmethoden inter- und intraindividuell Testleistungsunterschiede sowie Methoden zur Interpretation von Test*profilen* finden sich bei Lienert (1967, S. 454 ff.).

Schließlich müssen wir noch den seither etwas unkritisch verwendeten Terminus „Irrtumswahrscheinlichkeit“ genauer bestimmen. Der Begriff impliziert Rationale der Wahrscheinlichkeitstheorie und der (statistischen) Hypothesenprüfung. Bei diagnostischen Fragestellungen der vorgenannten Art gehen wir – auf Interpretationsebene – prinzipiell zwei Möglichkeiten von Risiken ein.

*Risiko erster Art:* Zufällige (inter- oder intraindividuelle) Ergebnisdifferenzen in einem Test werden als echte Unterschiede angesehen, d.h. die Testbefunde werden „überinterpretiert“. In diesem Falle begehen wir den *Fehler vom Typ  $\alpha$* : Wir verwerfen die Nullhypothese, obwohl sie (freilich nur mit einer Wahrscheinlichkeit von p = 32 % bzw. 16 % versus 5 % bzw. 2,5 % versus 1 % bzw. 0,5 %) <sup>44</sup> stimmt.

*Risiko zweiter Art:* Tatsächlich vorhandene Testleistungsunterschiede werden als zufallsbedingte Schwankungen gedeutet, d.h. die Testbefunde werden „unterinterpretiert“. In diesem Falle begehen wir den *Fehler vom Typ  $\beta$* : Wir akzeptieren die Nullhypothese, obwohl sie (freilich nur mit einer Wahrscheinlichkeit von p = 32 % bzw. 16 % versus 5 % bzw. 2,5 % versus 1 % bzw. 0,5 %) falsch ist. In der Experimentalpsychologie wird zumeist die Größe des Fehlers  $\alpha$  gewählt.

Beide Risikoarten oder Fehlertypen stehen in einem reziproken Verhältnis zueinander. Die allgemeine Meinung, daß das Testergebnis resp. eine bestimmte Hypothese um so besser abgesichert sei, je niedriger die Zufallswahrscheinlichkeit p für das Zutreffen der Nullhypothese, also das Signifikanzniveau, gewählt werde, ist somit nur bedingt richtig. Je rigoroser man ist, desto weniger wahrscheinlich ist es, daß die richtige Hypothese ver-

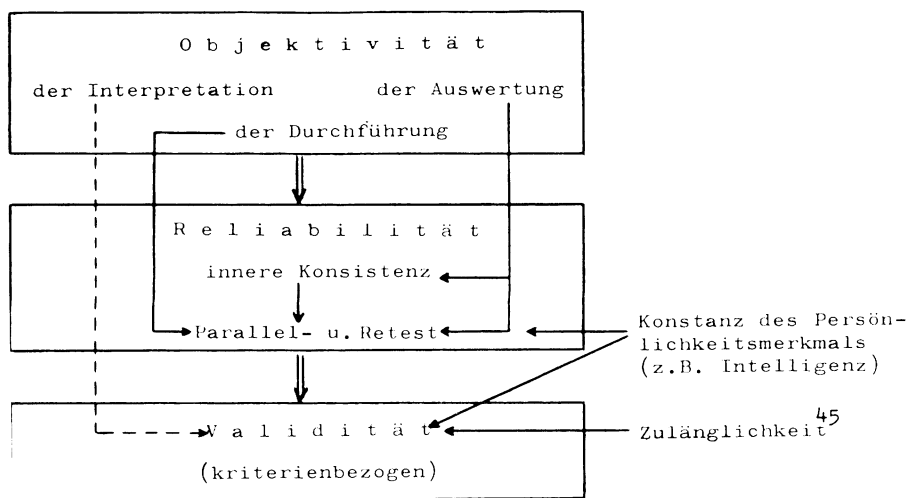
worfen wird; gleichzeitig ist aber die Wahrscheinlichkeit relativ hoch, eine falsche Hypothese zu akzeptieren. Die Faustregel „je rigoroser, desto besser!“ ist generell so nicht haltbar.

Praktisch wird die Favorisierung des Fehlertyps von der jeweiligen Fragestellung bzw. dem in der diagnostischen Problemsituation vermeintlich kleineren Übel abhängen. So wird man etwa in der Schul- und Studieneignungsermittlung (Begabungsdiagnostik) im allgemeinen bestrebt sein, den Fehler vom Typ  $\alpha$  kleiner (als den vom Typ  $\beta$ ) zu halten, ohne freilich allzu rigoros vorzugehen. In der Regel gibt man sich hier mit einem  $\alpha$ -Fehler-Risiko von 32 % (zweiseitiger Test) bzw. 16 % (einseitiger Test) oder auch – seltener – 5 % bzw. 2,5 % zufrieden; das Fehlerrisiko zweiter Art wäre natürlich um so höher, je niedriger das Signifikanzniveau für den Fehler erster Art angesetzt werden würde, d.h. konkret, wenn das Test-Normenband oder Konfidenzintervall am 5%-Signifikanzniveau statt am 32%-Signifikanzniveau orientiert wäre (vgl. dazu noch Feldt 1967). Weitere Beiträge zur Frage der Sicherung von Testwertdifferenzen werden wir im Rahmen der Behandlung des Normenproblems (Testeichung) und der Interpretation von Testbefunden (Gutachten) sowie im empirischen Teil dieses Werkes bringen.

### Validität

Während Meßfehler und Reliabilität über die *formale* Genauigkeit von Testwerten Auskunft geben, informiert die „Validität“ (engl. validity) darüber, „welche psychodiagnostischen Schlußfolgerungen die numerischen Resultate eines Tests zulassen und welchen Grad an Sicherheit solche Schlußfolgerungen aufweisen“ (Michel 1964, S. 47). Eine gängige Kurzformel lautet deshalb: *Validität ist die Genauigkeit, mit der ein Test das mißt, was er messen soll.* Im deutschen Sprachgebrauch wird in diesem Zusammenhang auch von *Gültigkeit*, von *Treffsicherheit*, von *diagnostischer Valenz* oder *Findewertigkeit* gesprochen.

Was den Zusammenhang zwischen Validität und Reliabilität betrifft, so gilt: Reliabilität ist die notwendige, aber nicht hinreichende Voraussetzung für die Validität. Die Wechselwirkung zwischen den drei Gütekriterien (Objektivität, Reliabilität, Validität) kommt in folgendem Schema, das wir Lienert (1967, S. 20 f.) entnommen haben, zum Ausdruck.



Lienert (loc. cit.) leitet hieraus folgende Regeln für die genannten Beziehungen ab:

- „1. Die Parallel- oder Retest-Reliabilität eines Testes kann nicht höher sein als seine Konsistenz oder seine Objektivität; ferner kann ein Test nicht valider sein als er reliabel ist.
2. Ein Test mit einer hohen Validität muß notwendigerweise auch hohe Objektivität, Konsistenz und Zulänglichkeit besitzen. Die Feststellung einer hohen Validität entbindet somit in gewissem Maße von der Überprüfung der übrigen Gütekriterien.
3. Ein Test mit einer ausreichenden Validität und einer geringen Konsistenz muß eine hohe Zulänglichkeit besitzen. Solch ein Test hat ausgezeichnete Verbesserungschancen, da man die Konsistenz testtechnisch im allgemeinen leicht erhöhen kann (z.B. durch Testverlängerung, d.h. Erhöhung der Aufgabenzahl; d. Verf.), während man gegenüber einem Mangel an Zulänglichkeit machtlos ist. Ein Test mit geringer Reliabilität eignet sich bestenfalls zum Vergleich von Gruppen hinsichtlich des zu untersuchenden Merkmals – hierfür ist eine Reliabilität von  $r_{tt} \approx 0.5$  ausreichend –, nicht hingegen für eine individuelle Differenzierung.
4. Ein Test mit einer geringen Validität und einer hohen Reliabilität muß notwendigerweise eine geringe Zulänglichkeit haben. Ohne inhaltliche Umarbeitung gibt es für diesen Test keine Verbesserungsmöglichkeit. Er eignet sich zwar gut zur Differenzierung von Individuen, jedoch hat man keine Gewißheit darüber, aufgrund welchen Persönlichkeitsmerkmals diese Differenzierung erfolgt.
5. Ein Test mit geringer Validität und geringer Reliabilität ist praktisch wertlos.“

Eine Vielzahl von Validitätskonzepten und Validierungstechniken, besonders im angloamerikanischen Raum, hat lange Zeit eher zur Verwirrung denn zur Klärung des wichtigsten Testgütekriteriums beigetragen. In den letzten Jahren setzte sich jedoch international eine 1954 von der APA (American Psychological Association) vorgeschlagene Einteilung in folgende *Validitätsarten* durch: Content Validity, Concurrent bzw. Predictive Validity und Construct Validity.

Bei Testvalidierungen sind prinzipiell zwei Schlußweisen relevant:

1. Schluß vom Testverhalten des Probanden auf sein Verhalten außerhalb der Testsituation mit den Varianten
  - a) *Repräsentationsschluß*: Direkter Schluß auf das Verhalten außerhalb. Dieser ist bei der Inhaltsvalidierung („Content Validity“) einschlägig.
  - b) *Korrelationsschluß*: Dieser ist für die Übereinstimmungsvalidität („Concurrent Validity“) und die Vorhersagevalidität („Predictive Validity“) bedeutsam.
2. Schluß vom Testverhalten auf „Dispositionen“, „Fähigkeiten“, „Eigenschaften“ u.ä., die als grundlegend für eben dieses Testverhalten (hypothetisch) angenommen werden, weshalb man in diesem Zusammenhang von (Gedanken-)Konstrukten spricht. Einschlägig ist hier die sog. Konstruktvalidität („Construct Validity“).

Entsprechend dem Einteilungsvorschlag der APA werden nun folgende *Validierungsverfahren* bzw. Validitätskategorien unterschieden:

1. *Inhaltsvalidität*. Hier wird das Testverhalten als repräsentatives Sample eines bestimmten Verhaltensuniversums aufgefaßt. Lienert spricht auch von „logischer“ oder „trivialer“ Validität (Gültigkeit), weil der Repräsentationsschluß evident ist. So gestattet etwa das Ergebnis eines Schulleistungstests für Rechnen, in dem best. Rechenaufgaben (z.B. Textrechnungen) vorkommen, den Schluß, daß der Proband Rechenaufgaben ähnlicher Art auch außerhalb der Testsituation gleich gut lösen wird. Ebenso kann man von einem Rechtschreibtest annehmen, daß er (in erster Linie) Rechtschreibkenntnisse prüft.

Der Grad der Inhaltsvalidität ist im allgemeinen nicht numerisch zu bestimmen. Er muß aufgrund logischer und fachlicher Überlegungen (Rating von Experten) allgemein ak-

<sup>45</sup> Mit „Zulänglichkeit“ bezeichnet Lienert den Grad dessen, was an „Gemeinsamkeit“ durch einen Test und sein Validitätskriterium (*Außenkriterium*) erfaßt wird. Der Begriff „kriterienbezogene Validität“ wird auf Seite 74 erläutert.

zeptiert werden. Bei Interpretationen via Repräsentationsschluß muß man sich freilich immer vor der *Gefahr der Überinterpretation* hüten. Beispielsweise bedeutet das Beherrschen von Vokabeln noch keine Sprachgewandtheit oder besondere Fremdsprachenbegabung. Bei guten Intelligenztests wird man deshalb in der Regel auf die Inhaltsvalidierung verzichten zugunsten der Kriterienvalidierung (vgl. Pkt. 2).

**2. Kriterienbezogene Validität.** Gemeint ist hier die Validität als Korrelation mit einem sog. Außenkriterium. Das Konzept entspricht der „klassischen“ Testtheorie, nach der Validität schlechthin als „Korrelation mit dem Kriterium“ bezeichnet wird (vgl. Gulliksen 1950). Der Ansatz ist einfach: Es wird nach der Enge des Zusammenhangs zwischen Testverhalten und jetzigem (*Übereinstimmungsvalidität*) oder zukünftigem (*Vorhersagevalidität*) Verhalten des Probanden außerhalb der Testsituation gefragt. Diesem Vorgehen liegt eine *empirisch* nachweisbare Wenn-Dann-Beziehung zugrunde, die keineswegs evident oder wissenschaftlich aufgeklärt zu sein braucht (Michel 1964). Das Bezugskriterium muß ausreichende Reliabilität und Validität haben. Desgleichen wird vom Test selber hinreichende Reliabilität (siehe oben) gefordert.

Für die Zusammenhangsberechnung wird meistens die Produkt-Moment-Korrelation nach Pearson oder eine daraus abgeleitete Formel benutzt. Der so ermittelte Korrelationskoeffizient ( $r$ ) drückt den Zusammenhang zwischen Test (test) und Kriterium (criterion) aus und symbolisiert als *Validitätskoeffizient* ( $r_{tc}$ ) den Grad der – kriterienbezogenen – Gültigkeit eines Tests. Da in diese Validität immer auch die (Un-)Reliabilität des Außenkriteriums (wie des Tests selber) eingeht, verwendet man häufiger sog. *Minderungskorrekturen*, die zu Lasten des Kriteriums gehende Unzuverlässigkeiten neutralisieren. Die kriterienkorrigierte Formel hierfür lautet

$$\text{crit. corr. } r_{tc} = \frac{r_{tc}}{\sqrt{r_{cc}}}, \text{ wobei } r_{cc} \text{ den}$$

Reliabilitätskoeffizienten des Validitäts- oder Außenkriteriums symbolisiert. Manchmal werden auch *doppelte* Minderungskorrekturen (crit. et test-corr.  $r_{tc}$ ) durchgeführt. Bei der Interpretation von Validitätskoeffizienten ist deshalb darauf zu achten, ob diese unkorrigiert, einfachkorrigiert (d.h. die Unreliabilität des Kriteriums berücksichtigend) oder gar doppeltkorrigiert (d.h. die Unreliabilität des Kriteriums *und* des Tests berücksichtigend) gewonnen wurden. Es leuchtet ohne weiteres ein, daß minderungskorrigierte Validitätskoeffizienten, z.B. von Intelligenztests, höher liegen müssen – etwa zwischen 0.6 und 0.8 – als unkorrigierte Validitätskoeffizienten, die selten Werte um 0.6 übersteigen (vgl. noch Lienert 1967, S. 280 ff. sowie Löwe & Heller 1972, S. 46 f.).

Bei der Validierung eines neuen Intelligenztests verwendet man häufig einen bewährten und gut standardisierten alten Intelligenztest als Bezugskriterium. Streng genommen ist damit aber nur geprüft, ob das neue Testverfahren ein Äquivalent des alten darstellt, d.h. dieses – vielleicht wegen sonstiger Vorzüge, etwa zeitlich-ökonomischer oder rationeller Gesichtspunkte – ersetzen kann. Umgekehrt sprechen negative Validierungsbefunde dieser Art nicht ohne weiteres gegen den neuen Test, der u.U. andere Faktoren der Intelligenz (s.S. 14ff.) erfaßt. In der modernen Testkonstruktion wird deshalb zunehmend häufiger die *Faktorenanalyse* als Hilfe der (Intelligenz-)Testvalidierung – zusätzlich – eingesetzt, zumal „faktorierte“ Tests vielerlei Vorteile für die Befundinterpretation, operationalisierte Klassifikationsansätze u.dgl.m. bieten (vgl. Heller 1970a, S. 107 ff.; Allinger & Heller 1972).

Schul- und sonderpädagogisch relevanter als die Übereinstimmungsvalidität ist die *Vorhersagevalidität* eines Tests, also die Frage, mit welcher Berechtigung man beispielsweise vom aktuellen Intelligenztest-Leistungsverhalten auf künftige Schul- und Bildungserfolge schließen kann. Damit wäre die wichtige Frage der *Schuleignungsermittlung* bzw. *Schulleistungsprognose* durch (Intelligenz-)Tests thematisiert. Das Außen- oder Validitätskriterium steht diesmal erst zu einem späteren Zeitpunkt, etwa in Form von Lehrerurteilen über den schulischen Bildungserfolg (Zensuren u.ä.), zur Verfügung. Bewährungskontrollen von Test-

aussagen dieser Art und Prognosen überhaupt sind freilich in weit stärkerem Maße von der Reliabilität versus Unreliabilität der Bezugskriterien beeinflusst als entsprechende Kontrollen (intelligenz)testdiagnostischer Aussagen, also Kontrollen über Übereinstimmungsvalidität. Ausführlicher werden wir die angeschnittenen Probleme noch im Zusammenhang mit eigenen empirischen Untersuchungsergebnissen erörtern (s.S. 161ff.).

3. **Konstruktvalidität.** Diese Methode ist die einzig mögliche beim Versuch, sog. projektive Testverfahren (z.B. Rorschach-Test, Thematischer Apperzeptions-Test) zu validieren. Häufig ist man auch bei der Gültigkeitsbestimmung anderer Verfahren, etwa Persönlichkeitsfragebogen oder „neuer“ Intelligenztests, d.h. Tests ohne Vorbilder, z.B. Kreativitätstests, auf die Konstruktvalidierung angewiesen.

Während bei der kriterienbezogenen Validität vorab der Zusammenhang zwischen Testergebnis und Kriterium interessiert, ist für die Konstruktvalidität eine (explizite) Theorie unerlässlich. „Die Konstruktvalidität zielt direkt ab auf die psychologische Analyse der einem Test zugrunde liegenden Eigenschaften und Fähigkeiten, also auf Beschreibungsmerkmale, die nicht in eindeutiger Weise operational erfassbar, sondern theoretischen Charakter haben, wobei freilich eine empirische Basis gegeben ist“ (Lienert 1967, S. 17). M.a.W.: In der Konstruktvalidierung wird eine Theorie experimentell bzw. empirisch überprüft. Dabei wird folgendermaßen vorgegangen. Zunächst wird das Konstrukt in ein theoretisches Bezugssystem eingeordnet, aus dem empirisch verifizierbare Hypothesen abgeleitet werden können. Konstrukte werden also hier „nicht auf Beobachtbares reduziert (wie es z.B. bei einer operationalen Definition geschieht), sondern so in ein „nomological network“ hineingebaut, daß Voraussagen – und zwar exakt prüfbare Voraussagen – über Beobachtbares möglich werden“ (Hörmann 1961, S. 49). Sodann werden Modifikationen nach größerer Genauigkeit und Stimmigkeit angestellt, wodurch eine optimalere Einordnung ins Bezugssystem erreicht werden soll. Hörmann bezeichnet diesen Vorgang als „Prozeß der sukzessiven Approximation“. Anschließend werden die klassischen Validierungstechniken (siehe oben) eingesetzt; vgl. noch Michel 1964, S. 57 ff. Das im folgenden geschilderte (fiktive) Beispiel einer Konstruktvalidierung, das wir Ingenkamp (1971, S. 68 f.) entnehmen, möge die Zusammenhänge noch einmal veranschaulichen.

„Man will z.B. Kreativität mit einem Test erfassen, es gibt in Deutschland kein ähnliches Verfahren und kein befriedigendes Kriterium, nach dem man Kreativität in bestimmten Situationen beurteilen könnte. Der Testautor entwickelt in Auseinandersetzung mit der Fachliteratur ein theoretisches Modell der Kreativität, beschreibt, welche Verhaltensweisen zu dieser Kreativität gehören, welche sich mit ihr nicht vereinbaren lassen und welche nicht relevant für Kreativität sind. Dann kann er Testaufgaben entwerfen und in einer Reihe von Untersuchungen klären, ob sich die theoretisch erwarteten positiven oder negativen Zusammenhänge zu einzelnen Verhaltensweisen tatsächlich nachweisen lassen. Wenn das gelingt, hat er für sein Modell, sein Konstrukt, Gültigkeit nachgewiesen. Ob das theoretische Modell dann auch allgemeinere Anerkennung findet, hängt von logischer Stimmigkeit der Theorie und der methodischen Anlage der Untersuchungen ab. Deutlicher als bisher ist hier vielleicht klar geworden, daß ein Test nicht ‚die‘ Kreativität oder ‚die‘ Intelligenz mißt, sondern ein Verhalten, das nach einer bestimmten Theorie Rückschlüsse auf etwas erlaubt, was nach dieser Theorie Kreativität oder Intelligenz genannt werden könnte.“

Damit wollen wir die etwas umfangreicher geratenen Ausführungen über die Testgütekriterien beenden und kommen nun zum letzten Punkt unserer testtheoretischen Einführung, zum Problem der Testnormen bzw. Testeichung.

#### d) Das Normenproblem (die Testeichung)

Für jeden brauchbaren Test müssen Normdaten zur Verfügung stehen, weil erst diese den Vergleich individueller Testleistungen mit der Gruppenleistung, d.h. der Testleistung der gewünschten Bezugsgruppe (Referenzpopulation), gestatten. Die Gewinnung von Test-

normen bezeichnet man auch als (Test-)Eichung. In der intelligenzdiagnostischen Praxis sind oft mehrfache Gruppenbezüge und somit variierte Normierungsgesichtspunkte (z.B. Altersnormen, Schultypnormen, Klassennormen, geschlechtsspezifische u.a. Normen) interessant. Der Autor eines Intelligenztests kann jedoch – aus ökonomischen Gründen – meist nur eine begrenzte Anzahl von Normdaten liefern; je nach Problemspezifität der Untersuchungsziele (Schul- und Studieneignungsermittlung, Lernbehinderten- oder Sinnesgeschädigtendiagnostik, Erfassung unterrichtlicher Kognitionseffekte, wissenschaftliche Begleitung und Kontrolle von Schulmodellversuchen u.ä.) sind die potentiellen Benutzer des betr. Tests über die im Testmanual angebotenen Testnormen hinaus auf die Gewinnung eigener Normen, d.h. den speziellen Bedürfnissen Rechnung tragender Testinformationen, angewiesen. Diese Prozedur ist meist sehr aufwendig und zeitraubend, weshalb gerade in spezifischen testdiagnostischen Fragestellungen oft entsprechende Informationsdaten fehlen. Diesem Mangel in einigen Teilgebieten intelligenzdiagnostischer Anwendungspraxis abzuhelpen, ist das Hauptmotiv der im empirischen Teil dieses Werkes mitgeteilten Forschungsbefunde (vgl. auch Tabellen-Anhang).

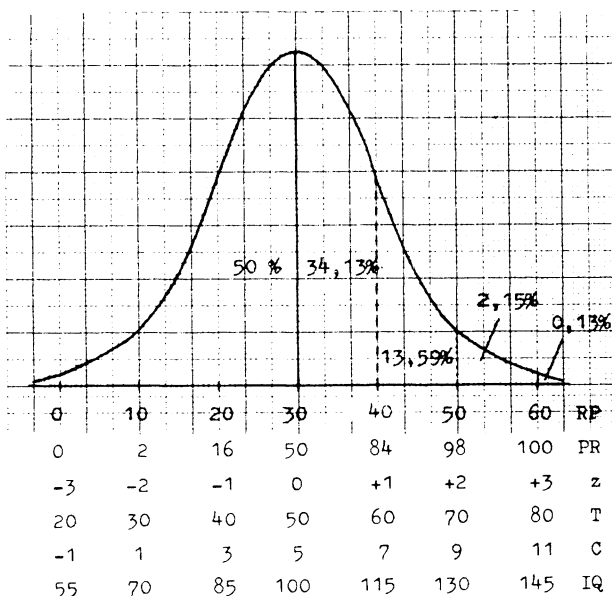
In der Regel wird man nicht die Gesamtpopulation, sondern nur eine Zufalls- oder repräsentative *Eichstichprobe* zwecks Normengewinnung testen können. Die Aussagekraft der so gewonnenen Testnormen hängt demnach von der Güte der Stichprobenbildung ab, weshalb für die Sample-Auswahl (Sampling) größte Sorgfalt gefordert und entsprechende Repräsentationsdaten (relevanter Merkmalsanteile oder "Strata") publiziert werden müssen. Je genauer die Strata der Referenzpopulation in der Eichstichprobe abgebildet sind, desto besser ist die diagnostische Valenz von Testnormen zu veranschlagen. Umgekehrt bedingen Mängel in der repräsentativen Zusammensetzung der Eichpopulation oder auch zu kleiner Stichprobenumfang eine mehr oder minder große Einengung der Normengültigkeit. Solche Mängel sind freilich nicht immer auszuschließen, da man mitunter (etwa in der klinischen oder sonderpädagogischen Diagnostik) auf extrem kleine Pbn-Mengen – selbst der entsprechenden Gruppenkollektive – angewiesen ist. Stratifikationskriterien i.S. einer *relativen* Repräsentanz (einschlägiger Kennmerkmale der Untersuchungsgruppe) gewinnen hier vorrangige Bedeutung und sind oft die einzige Möglichkeit, den Stichprobenfehler möglichst klein zu halten.

Bei den meisten Intelligenztests wird dem Probanden für jede richtige Lösung einer (beliebigen) Aufgabe 1 Pluspunkt gegeben; manchmal werden auch Teillösungen bzw. pro Aufgabe mehrere Punkte – je nach Schwierigkeitsgrad oder Qualität der Aufgabenlösung – angerechnet, so z.B. im HAWIK. Entsprechend der vorgeschriebenen Punkte-Auszählung (*Rohwertermittlung*) erhält dann der Proband Summenscores für die Leistung in den einzelnen Untertests oder/und im Gesamttest, womit auch schon der erste Schritt einer Testauswertung vollzogen ist.

Die *Rohwertscores* bringen die *absolute* Testleistung eines Pb oder einer Pbn-Gruppe zum Ausdruck. In der testdiagnostischen Praxis kann man damit freilich wenig anfangen. Hier interessieren meist interindividuelle, manchmal auch intraindividuelle, Vergleichsmöglichkeiten, die nur auf der Basis vorgegebener (Gruppen-)Maßstäbe realisierbar sind, d.h. die Testrohwerte (Pb-Leistung im Test) müssen in *Prozent-Rangplätze* (PR) oder *Standardwerte* (z.B. IQ, T, C, z) transformiert werden (siehe Grafik 2). Erst dann können wir die *relative* Position eines Individuums innerhalb seiner Bezugsgruppe (Referenzpopulation) bestimmen, wodurch die eigentliche Funktion eines Tests definiert ist (s.S. 57f.).

Der Begriff des Messens beinhaltet immer den *Vergleich mit einem vorliegenden Maßstab*. Für das *Messen* im physikalischen Sinne gelten deshalb die folgenden „klassischen“ Anforderungskriterien einer Meßskala: a) festgelegte Einheiten, b) gleiche Intervalle, c) absoluter Nullpunkt. Nun kennen wir allerdings auch physikalische Skalen ohne absoluten Nullpunkt, z.B. die Temperaturskalen. Dieses Beispiel zeigt schon,

daß man *verschiedene Ebenen des Messens* unterscheiden muß. Die 1. (*unterste*) *Ebene des Messens* (Nominal- oder Klassifikationskala), die nur einem einzigen Kriterium – der Äquivalenz – genügt, und die 4. (*oberste*) *Ebene des Messens* (Verhältnis- oder Rationalskala), die allen vier klassischen Kriterien – der Äquivalenz, der Beziehung zwischen den Positionen, der Intervallkonstanz, dem absoluten Nullpunkt – genügt, spielen so gut wie keine Rolle in der *Intelligenz* testung. Einschlägig sind vielmehr hier die 2. *Ebene des Messens* (Rang- oder Ordinalskala) und die 3. *Ebene des Messens* (Intervallskala).



Grafik 2. Transformation einer Rohwertskala (Rohpunkte) in Prozent-Ränge (PR) und Standardwertskalen (z-Werte, T-Werte, Centil-Werte, Abweichungs-IQ) sowie Flächen-Prozente (Prozent der Fälle über den Achsenabschnitten der Standardabweichungen) unter der Normalkurve.

Die *Rangskala* genügt zwei Kriterien, dem der Äquivalenz und dem der Beziehung, z.B. Größer-als-Beziehung, Besser-als-Beziehung, Intelligenter-als-Beziehung u.ä. Viele Testskalen sind auf diesem Niveau angesiedelt, d.h. sie sagen etwas über die *Rang*-position des Individuums, nichts aber über die Größe der Rangunterschiede (Intervalle) aus. Solche Aussagen sind erst auf dem Niveau der *Intervallskala*, wo als drittes Kriterium die Intervallkonstanz hinzukommt, möglich. Damit ist das wichtigste Unterscheidungsmerkmal zwischen beiden Skalentypen (Rangskala versus Intervallskala) bezeichnet. So sind etwa Testleistungsunterschiede an Hand einer Wechsler-IQ-Skala immer direkt vergleichbar, d.h. *gleiche IQ-Abstände repräsentieren hier immer auch gleiche tatsächliche Intelligenzunterschiede*. Dies ist bei der PR-Skala nicht der Fall. Gleiche Rohwertunterschiede sind im Dichtemittel bzw. in der Nähe des Median durch große PR-Differenzen, an den Verteilungsenden der Perzentilskala durch sehr kleine PR-Differenzen repräsentiert. Man kann sich die %-Rangskala als ein an den beiden Enden überdehntes Gummiband vorstellen, wo sich die PR-Einheiten in der Mitte zusammendrängen, an den Enden aber unverhältnismäßig weit streuen. Auch



die Schulnoten (SN) sind häufig „nur“ Rangskalen, insofern die Schülerleistungen zwar rangmäßig richtig eingestuft sind, die einzelnen Rangabstufungen aber ungleichen Leistungsabständen der Schüler entsprechen können. So ist der Leistungsunterschied zwischen der Zensur 3 und 4 nicht ohne weiteres der gleiche wie der zwischen der Note 1 und 2 etc. Daraus folgt, daß bei Prozenträngen keine Mittelwerte (M) und Standardabweichungen ( $\sigma$  bzw. s), sondern lediglich der Median (Med) und Quartile (Q) berechnet werden dürfen. Die in Grafik 2 und Tabelle VIII (Anhang) vorgenommene Gegenüberstellung von Rang- und Standardwertskalen gilt deshalb auch nur unter der Voraussetzung der Normalverteilung.

1. **Standardnormen:** Diese sind nur brauchbar, wenn die Verteilung der empirisch ermittelten Test-Rohwerte der Eichstichprobe einer Gaußschen Kurve (Normalverteilung) entspricht. Unter dieser Voraussetzung können die Rohwerte X jederzeit in eine Standardskala mit z-Werten transformiert werden. Dies geschieht nach der Formel  $z = (X - M) / s$ . Grafik 2 veranschaulicht als Beispielfall die Rohwert-Transformation der PMT-Ergebnisse von 80 tauben Hauptschülern der 5. Klasse. Bei  $M = 29,9125 \sim 30$  RP und  $s = 10,2770 \sim 10$  RP ergaben sich für die einzelnen (möglichen) RP-Werte von 1 bis 60 über die z-Wert-Berechnung (siehe oben) die in Tabelle Ib (Anhang) aufgeführten IQ-Punkte für die 5. Hauptschulklasse, d.h. deren PMT-Normen. Durch lineare Transformation kann nämlich die – wegen der Vorzeichen und Kommastellen – unpraktische z-Wert-Skala in eine beliebige Standardwertskala (mit ganzen Zahlen bzw. positiven Werten) überführt werden. Durch eine Skalentransformation dieser Art wird freilich keine meßtheoretisch höherwertige Skala gewonnen. Jede Standardwertskala ist vielmehr durch M und s definiert, wie eine kurze Auswahl gängiger Normenskalen (mit entsprechenden Testhinweisen) illustriert. Die Beispiele beziehen sich ausschließlich auf die im empirischen Teil (siehe unten) referierten Tests.

(Abweichungs-)IQ-Skala:	M = 100	s = 15 (HAWIK bzw. HAWIE, PMT)
T-Skala:	M = 50	s = 10 (FTU 4-6)
C-Skala:	M = 5	s = 2 (LPS bzw. PSB)
WP-Skala:	M = 10	s = 3 (HAWIK-Subtests)

2. **Standardnormen-Äquivalente:** Um auch bei anormalen Verteilungen Rohwerte in Standardwerte überführen zu können, was durch lineare Transformation nicht erlaubt ist (siehe obige Kautel!), werden die Rohwerte zuweilen durch *Flächentransformation* einer Normalverteilung subsumiert (z.B. Orlik 1967). Das zuerst von McCall entwickelte und von ihm zu Ehren Termans als T-Norm bezeichnete Verfahren entspricht bei tatsächlich gegebener Normalverteilung der Rohwerte der oben beschriebenen linearen Transformation, weshalb wir die Parameter ( $M_T = 50$  und  $s_T = 10$ ) bei den Standardwertskalen mitaufgeführt haben. Die im Rahmen unserer Testuntersuchungen nicht verwendete, sehr aufwendige Prozedur der *Normalisierung* hat Lienert (1967, S. 339 ff.) ausführlich dargestellt.
3. **Prozentrangnormen:** Sie sind theoretisch unproblematisch und auch praktisch leicht zu ermitteln durch fortlaufende Summation der durch die Rohwerte repräsentierten Flächenstücke. Hierbei werden die Teilflächen jeweils in Prozenten der Gesamtfläche angegeben (siehe Grafik 2).

PR-Normen sind verteilungsfreie Normen. Sie setzen nur Messung auf dem Niveau der Ordinalskalen voraus und sind deshalb gerade bei Schultests, wo häufig die Anforderungskriterien von Intervallskalen nicht erfüllt werden, sehr beliebt. Der WST, BIT, PIT sowie eine Reihe weiterer Tests, auf die wir nachher noch zu sprechen kommen, bedienen sich beispielsweise der Prozentrangnormen. Sofern man die unterschiedliche Dichtedifferenzierung – Nachteil aller Perzentilskalen! – bei der Befundinterpretation beachtet, sind PR-Normen sehr brauchbar, nicht zuletzt wegen ihrer starken An-

schaulichkeit für den Interpretieren. So bedeutet etwa ein PR von 10 in einem Test, daß der betr. Proband besser als 10 % und schlechter als 90 % seiner Bezugsgruppe (Altersgruppe, Schulgruppe, Klassengruppe usw.) ist, während der 98. PR aussagt, daß der betr. Proband 98 % seiner Referenzpopulation in der gemessenen (Intelligenz-) Dimension übertrifft. Ein Blick in die im Anhang beigelegte Transformationstabelle von Testnormen (Tabelle VIII) oder das Schaubild von Grafik 2 informiert uns weiterhin, daß dieser Rangplatz (97,5 PR) den Standardwerten 2 z, 70 T, 9 C, 130 IQ bzw. 16 WP (in den HAWIK-Subtests) versus den Schulzensuren 1 (5stufige Skala) bzw. 1/2 (6stufige Skala) entspricht. Mit Hilfe der Transformationstabelle sind also – ohne Rechenarbeit – die in verschiedenen Normenskalen ausgedrückten Testresultate unterschiedlicher Provenienz direkt miteinander vergleichbar, sofern normale oder annähernd symmetrische Datenverteilungen (im letzteren Falle beschränkt sich die Vergleichbarkeit auf den mittleren Skalenbereich) vorliegen. Andernfalls wären Rangplätze nur über den Weg der Normierung (Methode nach McCall) *direkt* mit Standardwertskalen vergleichbar. Siehe hierzu Tab. IX im Anhang.

4. *Äquivalentnormen*: Diese Normenarten gehen alle auf das Intelligenz-Alter (IA) von Binet und Simon (1908) zurück. „Das Intelligenzalter bezeichnet das Niveau der intellektuellen Entwicklung, ausgedrückt als Äquivalent desjenigen Lebensalters, in dem im Durchschnitt dieses Niveau erreicht wird. Analog zum IA werden – vor allem in den USA – andere Äquivalentnormen, wie Entwicklungsalter, „educational age“ und „grade norms“ verwendet“ (Michel 1964, S. 31).

Meßtheoretisch und psychologisch ist diese Art von Normen jedoch sehr fragwürdig. Ein Zehnjähriger mit IA = 7 hat eben *nicht* die Intelligenz eines Siebenjährigen, sondern die eines minderbegabten Zehnjährigen! Im Hinblick auf den Schrumpfungsprozess, dem IA-Einheiten mit zunehmendem Lebens-Alter (LA) unterliegen, schlug W. Stern bereits 1912 als neues Intelligenzmaß den sog. Intelligenz-Quotienten (IQ) vor, in dem IA und LA aufeinanderbezogen werden:

$$(\text{Sternscher}) \text{ IQ} = \frac{\text{IA}}{\text{LA}} \cdot 100$$

Um Dezimalstellen zu vermeiden, wird der Ausdruck gewöhnlich mit der Konstante 100 multipliziert.

Der Sternsche IQ war lange Zeit das berühmteste Intelligenzmaß überhaupt. Seine Problematik liegt jedoch darin, daß die Streuung des „Intelligenzalters“ *nicht* direkt proportional zum Lebensalter wächst, weshalb Intelligenz-*Quotienten* aus verschiedenen Altersstufen nur bedingt, nämlich nur bei konstanter IQ-Variabilität einzelner (interessierender) Altersstufen, miteinander verglichen werden können. Heute ist deshalb fast durchweg der erstmals von D. Wechsler eingeführte, sogenannte *Abweichungs-IQ* in Verwendung. Der Abweichungs- oder Wechsler-IQ unterscheidet sich vom Sternschen IQ dadurch, daß er den Leistungsabstand eines Individuums vom Mittelwert seiner Bezugsgruppe angibt, also durch die „Abweichung“ vom arithmetischen Mittelwert – in Sigma-Einheiten der Standardabweichung ausgedrückt – die *relative Position* eines Probanden innerhalb seiner Altersgruppe (beispielsweise) bestimmt. Im Gegensatz zum Sternschen IQ wird hier nicht mehr das „Intelligenzalter“ auf das Lebensalter des Pb bezogen, d.h. überhaupt kein Quotient mehr gebildet. Der *Abweichungs-IQ* ist wie jede andere Standardwertskala (Standardnorm) definiert durch Mittelwert (M = 100 IQ) und Standardabweichung (s = 15 IQ). Wenn künftig – ohne Zusatz – von IQn die Rede ist, so ist ausschließlich dieser Abweichungs-IQ gemeint (vgl. auch Punkt 1 oben).

Bei den modernen Intelligenztests werden durchweg Normmaßstäbe auf Intervallskalenniveau oder Prozentränge angeboten. Diese beziehen sich allerdings meist auf un-

ausgelesene oder ausgelesene größere (vollständige) Schülerpopulationen. Dagegen besteht nach wie vor in der sonderpädagogischen Diagnostik, vorab bei Sinnesgeschädigten (sensorisch deprivierten Personen), ein erheblicher Mangel an populationspezifischen Normen sonst mehr oder weniger brauchbarer Intelligenztests. Unsere im Anhang (Tabellen I bis VII) beigefügten *Testnormen für Hör-/Sprachbehinderte* mögen diese Lücke ein wenig schließen helfen. Mit Rücksicht auf die geringen Pbn-Zahlen in manchen Teilstichproben und/oder meßtheoretische Kautelen in bezug auf einige (Fehler-)Variablen wurden in den Tabellen IIb, V, VIa, VIb und VII die in der Intelligenztestdiagnostik relativ selten verwendeten *Großnormen* (Median- und Quartilnormen) aufgeführt. An Hand dieser Normdaten können die getesteten Schüler (lediglich) in 4 Leistungsgruppen eingeteilt werden: 1. – 24. PR, 25. – 49. PR, 50. – 74. PR, 75. – 100. PR. In der pädagogischen Diagnostik genügt diese grobe Klassifikation vielfach – zumindest im Hinblick auf die in den betr. Testdimensionen erfaßten Persönlichkeits-/Intelligenzmerkmale – den praktischen Erfordernissen.

Am Schluß dieses Kapitels sei ergänzend zu der bereits erwähnten Testliteratur noch auf folgende Veröffentlichungen aufmerksam gemacht: Anastasi (1968), Belser (1967), Cronbach (1970), Gulliksen (1950), Ingenkamp & Marsolek (1968), Klix et al. (1967), Priester (1958), Rösler et al. (1970), Wechsler (1961), Wewetzer (1964, 1972), Witzlack (1971). In den Schulbibliotheken finden sich häufig noch Ingenkamp (1963) oder Samstag & Baus (1962), denen aber heute teilweise nur noch historische Bedeutung zukommt.

### III. Schülerbeurteilung und Bildungsberatung

#### 1. Ziele und Aufgaben der Intelligenztestanwendung in Schule und Sonderpädagogik

Psychologische Tests sind nie Selbstzweck, sondern psychodiagnostische Hilfen in wichtigen menschlichen Entscheidungssituationen. *Intelligenz* tests in der Bildungsberatung (Schule, Sonderpädagogik u.ä.) sollen z.B. (schul-)pädagogische Fragen, in denen intelligenzrelevante Probleme zur Entscheidung anstehen, klären helfen. Hierbei treten zwei Problemkomplexe in den Vordergrund: Erstens die Anwendung von Intelligenztests zum Zwecke der *Begabungs- oder (Schul-)Eignungsdiagnose*. Pädagogische „Auslese“-Urteile – *selection* nach Cronbach & Gleser – sind am Ende der Grundschule (Übergang von der Primar- zur Sekundarstufe), beim Schulabschluß der Haupt- oder Realschule (Berufsfindung versus Schulsystemwechsel, z.B. Übertritt von der Hauptschule zur Realschule oder von der Realschule zum Gymnasium – seltener umgekehrt) und auf der Primarstufe des Gymnasiums (Ermittlung der Studieneignung bzw. Berufseignung) aktuell. Die meisten testpsychologischen Eignungsuntersuchungen – in der BRD jährlich schätzungsweise über 1 Million – finden gegenwärtig im Rahmen der Berufsberatung statt. Da diese durch die psychologischen Dienste der Arbeitsämter, die das Monopol der Berufsberatung in Deutschland beanspruchen, besorgt werden, brauchen wir auf diese Probleme nur am Rande (z.B. bei der Erörterung sonderpädagogischer Fragen) einzugehen. Im Kontext *pädagogischer* Diagnostik werden unter dem erstgenannten Punkt vor allem Probleme der Schul- und Studieneignungsprognose bedeutsam. So werden Entscheidungen über den optimalen Bildungsgang eines Schülers gefordert, worauf die *Schullaufbahnberatung* Auskunft geben soll. Hierzu gehört auch die frühzeitige Erkennung der Notwendigkeit bzw. Angemessenheit sonderpädagogischer Beschulung (z.B. Besuch der Lernbehindertenschule) und – im weiteren Sinne – die Ermittlung der sog. Schulreife, also der Schuleignung für die Primarstufe (Grundschule), wobei besonders Einschulungsprobleme virulent werden. Aber auch bei Übergängen zwischen verschiedenen Schulsystemen (Hauptschule, Realschule, Gymnasium) außerhalb der „Nahtstellen“, wie sie das Prinzip der Durchlässigkeit fordert, bei der Aufnahme von Bewerbern des Zweiten Bildungsweges u.ä. Einrichtungen werden laufend Intelligenztests als Entscheidungshilfen benötigt. Fast immer geht es hierbei um die Voraussage schulischer Bewährung, d.h. um Schuleignungs*prognosen* (genauer: *Schulleistungsprognosen*; vgl. S. 58f.). Bereits in den historischen Anfängen des Testeinsatzes in Schulen ging es um solche Ausleseprobleme. Binet und Simon (1908) in Frankreich (Selektion Schwachsinniger, d.h. sog. Hilfsschüler oder – wie wir heute sagen – lernbehinderter Sonderschüler), Stern, Bobertag, Hylla und Lämmermann in Deutschland (Begabtenauslese) wären hier wohl als bedeutendste Vertreter der Schul-(Test-)Psychologie zu nennen (vgl. Tent 1969, S. 28 ff.).

Der zweite Komplex, in dem Intelligenztests eine wichtige Entscheidungsfunktion übernehmen, betrifft im engeren Sinne Schulleistungsprobleme: a) *Probleme eines leistungsdifferenzierten Unterrichts*, z.B. innerhalb der Gesamtschule, der Förderstufe, der Hauptschule (A- und B-Kurse) oder schulklasseninterner Gruppierung – *placement* nach Cronbach & Gleser; b) *Schulleistungsanalyse*, z.B. bei individuellem Leistungsveragen (Underachievement, Legasthenie, affektiven oder anderen Leistungsstörungen), aber auch bei Leistungsbehinderungen qua sensorischer versus sozio-kultureller Deprivationserrscheinung. Im Gegensatz zu oben handelt es sich hier fast immer um den Vergleich der Intelligenz(testaussage) mit dem aktuellen, d.h. zum Zeitpunkt der Untersuchung beobachteten – nicht zukünftigen – Leistungsstand des betr. Probanden. Aufschlüsse über kognitive und nichtkognitive Voraussetzungen, z.B. persönlichkeitspsychologische (Motive, Interessen, Ängstlichkeit), sensorische, körperliche oder soziale Bedingungen intellektuellen Leistungsverhaltens, und deren mögliche Abgrenzung (Differentialdiagnose der

kognitiven und nichtkognitiven Bedingungen) sind ja in solchen Problemsituationen unabdingbare Voraussetzung für gezieltes pädagogisches Handeln.

„Handelt es sich um einen nur partiellen Ausfall, ist also die Intelligenz trotz schlechter Schulleistung durchschnittlich oder gut, so wird man sich bemühen, nach anderen Gründen für den Leistungsausfall zu suchen. Diese können im emotionalen Bereich liegen, es mag sich auch um Ausfälle eng umgrenzter kognitiver Funktionen handeln. Unter den Sorgenkindern, die den Pädagogen solche Fragen aufgeben, ist wohl das bekannteste das Grundschulkind, das eine Lese-Rechtschreibschwäche (= Legasthenie; d. Verf.) zeigt, gleichzeitig jedoch durchschnittliche oder sogar gute Intelligenzleistungen zustande bringt.

Mit Hilfe eines Intelligenztests, der in der beschriebenen Weise verwendet wird, läßt sich also klären, inwieweit es sich bei einem Schulversagen um einen allgemeinen oder um einen mehr spezifischen Leistungsausfall handelt. Eine solche Testanwendung könnte man, das Wort 'Analyse' in weitem Sinn gebrauchend, die ‚analytische‘, d.h. auf die Analyse von Schulleistungen gerichtete Anwendung von Intelligenztests nennen.

In der Regel ist, wie beschrieben, ein äußerer Anstoß, das Problem eines Leistungsveragens, notwendig, damit es zu einem derartigen Gebrauch von Intelligenztests kommt. Das ist wahrscheinlich die vernünftigste Praxis, und es soll hier keine Neuerung vorgeschlagen werden. Aber es soll darauf hingewiesen werden, daß die Logik dieses Vorgehens keineswegs die Beschränkung auf die Versager notwendig macht. Ähnlich wie bei den Versagern könnte man bei den Erfolgreichen und bei den Mittelmäßigen fragen, ob ihrem schulischen Leistungsstand ein ungefähr gleiches Intelligenzniveau (= Achievement; d. Verf.) entspricht oder ob zwischen schulischer Leistung und Intelligenz ein wesentlicher Unterschied (= Overachievement versus Underachievement; d. Verf. – vgl. auch S. 163 in diesem Buch sowie ausführlicher Heller 1969 a, S. 164 ff., 1969 b, S. 384 ff. und 1970a, S. 72) besteht. Allerdings, solchen Fragen fehlt das praktische Gewicht; so wenig wie von unserer Gesellschaft wird man von unserer Schule erwarten können, daß sie in einem Erfolg ebensoviel Problematik erkennt wie in einem Versagen“ (Wendeler 1970, S. 44).

Schließlich soll die Rolle der Intelligenztests bei der *wissenschaftlichen Begleitung* von Modellproben im Bereich der Schulpädagogik erwähnt werden. Kontrollen über die Effizienz unterschiedlicher Schulsysteme (verschiedene Formen der Gesamtschule versus dreigliedrige Sekundarschule), neuer Bildungseinrichtungen (z.B. Ganztageschule, Kollegstufe des Gymnasiums bzw. der Berufsschule) oder auch (didaktischer) Unterrichtsversuche sind ohne die Hilfe von Intelligenztests – in diesem Zusammenhang häufig als „Vortest“ versus „Nachttest“ eingesetzt – kaum möglich. Eine gute Information über die angesprochenen Probleme geben Raatz (1971) und ausführlicher Eigler et al. (1971). Neben mehr wissenschaftlichen Interessen (etwa der Bildungsforschung) geht es bei solchen Testkontrollen natürlich auch um individuelle Belange der betr. Schüler, also wieder um Analysen des Leistungsstandes und der Leistungsprognose, damit gezielte Förderungsmaßnahmen gegebenenfalls eingesetzt bzw. erprobt werden können. Praktisch ist hier wieder der gesamte Themenkatalog von Aufgabenzielen, wie wir ihn skizzierten, einschlägig. In der Sonderpädagogik treten zuweilen spezifische (bisher nicht genannte) Probleme hinzu, die wir jedoch aus didaktischen Gründen erst später bei der Erörterung unserer empirischen Untersuchungsbefunde (s.S. 185ff.) behandeln wollen. Zugleich sei noch auf die früheren Ausführungen zum Thema *Begabung und Bildungsförderung* (vgl. S. 45ff., bes. S. 51-54) verwiesen.

Wenn von (Intelligenz-)Tests und deren Einsatz in der Schulpraxis die Rede ist, so darf ein Problem nicht unerwähnt bleiben: der gegenwärtige Ausbildungsstand unserer Lehrer in pädagogischer Diagnostik. Dieser läßt – von Ausnahmen abgesehen – viele Wünsche offen und ist nicht erst in jüngster Zeit von Kennern der Situation beklagt worden. Als wesentliche Voraussetzungen für eine qualifizierte Testanwendung nennt beispielsweise Ingenkamp (1963, S. 22 f.) folgende *Anforderungskriterien an den Testleiter* (Lehrer):

1. Der Testleiter muß über *Grundkenntnisse der Testtheorie* und psychologischen Diagnostik verfügen. Detaillierter sollte er über die *Testgütekriterien* (Objektivität, Reliabilität, Validität) Bescheid wissen, um sowohl die richtige Testwahl treffen als auch den Wert der ge-

wählten Testverfahren beurteilen und mögliche Fehlerquellen oder Grenzen (der Anwendung bzw. diagnostischen Valenz) einschätzen zu können. Siehe unsere Ausführungen im II. Kapitel (S. 55ff.).

2. Der Testleiter muß über den Stand der wissenschaftlichen (psychologischen) Forschung auf dem Gebiet der *Untersuchungsdimension(en)* informiert sein. „Wer z.B. einen Intelligenztest anwenden will, sollte also mindestens in großen Zügen über die Psychologie der Intelligenz und die Problematik der Theorien und Begriffe Bescheid wissen.“ Eingehend wurden diese Fragen im I. Kapitel (S. 9-47) behandelt.

3. In engem Zusammenhang mit dieser Forderung muß für jeden Test- oder Untersuchungsleiter die Zielvorstellung existieren, die Bedeutung der untersuchten Eigenschaften und Fähigkeiten im Hinblick auf den gesamtpersonalen resp. sozialen und bildungspolitischen Kontext zu kennen und damit die Relevanz der testdiagnostischen Untersuchung überhaupt abschätzen zu können. „Wer mit einem Intelligenztest arbeitet, muß also wissen, welche Bedeutung die intellektuelle Leistungsfähigkeit im Persönlichkeitsgefüge hat, welche Anforderungen an sie z.B. beim Übergang auf die Oberschule gestellt werden.“ Letzten Endes geht es hier um die Frage „Wozu soll eine Testuntersuchung stattfinden?“ bzw. „Wem dient (nützt) die Untersuchung?“ – Siehe auch unsere Ausführungen auf S. 47ff.

4. Vertrautheit mit der Test- und Untersuchungstechnik, also der eigentlichen „Handhabung“ psychologischer Testmethoden, ist zwar eine notwendige, aber noch keine hinreichende Voraussetzung für eine effiziente (test-)diagnostische Untersuchung. Der Untersuchungsleiter muß auch in der Lage sein, die Erhebungsdaten (Testergebnisse, Fragebogen- und Beobachtungsdaten) sachgerecht auszuwerten resp. zu *interpretieren*, also die Befunde angemessen zu deuten. Über das „Wie“ wird im übernächsten Abschnitt zu handeln sein. Der geforderte Wissenskanon des Interpreters erstreckt sich von *meßtheoretischen Grundkenntnissen* (z.B. über den Meßfehler, die kritische Differenz zweier Testwerte oder Profildifferenzen) über Einsichten in den *probabilistischen* (wahrscheinlichkeitstheoretischen) *Charakter aller Testaussagen* (vgl. *Fehlerrisiko*) bis hin zu *impliziten* Persönlichkeits-, Intelligenz- und anderen einschlägigen *Theorien*, einschließlich bildungstheoretischen Modellen. Schließlich wäre als Spezifikum für die *pädagogische* Diagnostik noch zu nennen.

5. „Der Lehrer darf seine Schüler nur dann testen, wenn dazu eine *pädagogische Notwendigkeit besteht, wenn er zur Erfüllung seines Unterrichts- und Erziehungsauftrages durch den Test wesentliche Förderung erfährt*“ (Ingenkamp, a.a.O.).

Möglichkeiten und Grenzen des testdiagnostischen Einsatzes in der Schule und Sonderpädagogik sind somit formalinhaltlich abgesteckt. Im pädagogischen Interesse – des Schülers wie des Lehrers – gilt es nun, die aufgezeigten Möglichkeiten zu nutzen. Dieses Vorhaben zu erleichtern, ist eine der wichtigsten Intentionen dieses Buches. Bevor wir uns vollends dem empirisch-praktischen Teil zuwenden, seien noch zwei Problemkreise, auf die verschiedentlich schon hingewiesen wurde, angesprochen: die Vervollständigung des Erhebungsdesigns und das pädagogische Gutachten. Weitere Problemfragen, insbesondere soweit sie sich noch in der (intelligenz)testdiagnostischen Praxis für den Schulpädagogen stellen, sollen dann bei der Erörterung konkreter Lösungshilfen – im Rahmen unserer empirischen Testbefunde – behandelt werden.

## 2. Komplettierung der testdiagnostischen Erhebung (Design der Datenerfassung)

Die *testdiagnostische* Erhebung liefert in Problemsituationen der verschiedensten Art zu meist unentbehrliche, gleichwohl selten für eine treffsichere Entscheidung (anstehender Problemfragen) hinreichende Dateninformationen. Hierzu bedarf es ergänzender Erfassungsmethoden, wobei a) die Verhaltensbeobachtung, b) der Fragebogen und c) das psychodiagnostische Gespräch (Exploration und Anamnese) die größte Bedeutung erlan-

gen. Je nach *Art* und *Qualität* – manchmal auch *Anzahl* – der verwendeten *Tests* (z.B. vollstandardisierte Intelligenz- oder Leistungstests versus sog. projektive Tests) resp. der Besonderheit von Untersuchungs*ziel* (Intelligenz- oder Schulleistungsdiagnose versus Schuleignungsprognose oder Schulleistungsanalyse eines Problemfalls) und Untersuchungs*situation* (Gruppentestung versus Individualuntersuchung) wird der Einsatz der ersten oder/und zweiten Ergänzungsmethode genügen oder es werden alle drei Zusatzverfahren erforderlich. Zeitliche und ökonomische Gesichtspunkte sollten jedenfalls nicht primär über die Hereinnahme von Zusatzverfahren in die Methodenbatterie versus ihren Verzicht entscheiden. Im Rahmen schulischer Begabungsuntersuchungen kommt vor allem den beiden erstgenannten Methoden (Verhaltensbeobachtung<sup>46</sup> und Fragebogen) Gewicht zu, weshalb wir auf diese ausführlicher eingehen wollen. Exploration und Anamnese als Varianten des psychodiagnostischen Gesprächs stellen vorab in der klinischen Psychodiagnostik – und Therapie – sowie bei der Individualdiagnose bzw. Prognose des Einzelfalls sehr oft unentbehrliche Informationsquellen oder Medien dar.

#### a) Verhaltensbeobachtung

Auf die Verhaltensbeobachtung sollte bei keiner Testuntersuchung verzichtet werden. In der testdiagnostischen Praxis ist häufig nur die sog. *Gelegenheitsbeobachtung*, d.h. die – fallweise – Erfassung hervorstechender Verhaltensweisen einzelner Probanden, möglich. Mehr Informationen sowie verlässlichere und validere Auskünfte bietet allerdings die *systematische*, d.h. planmäßig veranstaltete, Verhaltensbeobachtung, wozu jedoch ein Testleiter nur bedingt in der Lage ist. Wegen der allgemein verhältnismäßig begrenzten individuellen Aufmerksamkeitsspanne bedarf es bei einem solchen Vorhaben in der Testsituation eigener, von der Person des Testleiters unabhängiger, Beobachter, die aus personellen und finanziellen Gründen freilich selten zur Verfügung stehen. Hingegen wird man im Rahmen wissenschaftlicher Testuntersuchungen auf systematische Beobachtungen dieser oder jener Form (siehe Fragebogen) nicht verzichten können.

Bei der Verhaltensbeobachtung – wie bei der Erlebnisbeobachtung – als Prozeß müssen prinzipiell drei (Abstraktions-)Stufen unterschieden werden, nämlich erstens die Beobachtung im engeren Sinne, zweitens die Beschreibung oder Deskription der beobachteten Phänomene und drittens die Beurteilung oder Deutung der deskribierten Tatbestände = Interpretationsphase. Unter *Beobachtung* in diesem eingeschränkten Sinne ist nicht bloße Wahrnehmung gemeint. Beobachtung ist immer fixierend auf bestimmte Phänomene resp. Phänomenauschnitte gerichtet, d.h. zentriert und selektierend zugleich, insofern sie stets etwas Bestimmtes ausmachen will. M.a.W.: Bei jeder Beobachtung besteht eine gewisse *Absicht*, also der Wunsch, von jeweils uninteressanten Verhaltensweisen (des Probanden) „abzusehen“, das Verhalten nur in bestimmten Hinsichten zu beobachten. Die Selektivität der Beobachtung stellt jedoch kein grundsätzliches Problem der Observational Techniques (der Name steht hier *pars pro toto*) dar. Das wissenschaftsmethodische Problem der Beobachtungstechniken liegt in der Deskription, genauer: in der schieflichen Trennung von Beschreibung und Beurteilung, also auf der Stufe der Verbalisierung. Graumann (1964, S. 90 ff.) hat in diesem Zusammenhang folgende vier *Beschreibungsmodi* herausgearbeitet:

1. Das *verbale* Niveau. Hiermit ist eine „reine“ Beschreibung prozessualer Abläufe gemeint. Dieser Beschreibungsstil ist arm an Adverbien, d.h. Qualifikationen werden weitgehend vermieden. Beispielfhaft seien etwa folgende Protokollsätze genannt: „Pb A faßt wiederholt (mehrmals, fortlaufend, viermal in 5 Minuten, usw.) nach dem Arm der Testleiterin.“

---

<sup>46</sup> Die *Erlebnisbeobachtung* (die sog. Intro- bzw. Retrospektion) spielt im Rahmen schulischer Testerhebungen nur gelegentlich eine Rolle (vgl. z.B. Heller 1972 b, S. 51).

– „Pb B stellt immerzu Fragen, auch nach wiederholter Instruktion der einzelnen Testaufgaben bzw. kontrolliertem Instruktionsverständnis.“ – „Pb C kaut ständig an seinem Füllfederhalter. Er spricht ständig mit dem Nachbarn. Er wischt sich den Schweiß von der Stirn. Seine Augen tränen. Sein Gesicht läuft öfters rot an.“

2. Das *adverbiale* Niveau. Hiermit sind bereits erste Qualifikationen der Abläufe erfaßt, wie etwa folgende Beispiele verdeutlichen: „Pb A greift *zitternd* nach dem Arm der Testleiterin.“ – „Pb B stellt *unangefochten* durch die Zurechtweisung des Testleiters fortgesetzt seine Fragen.“ – Einen Schritt weiter geht etwa diese Beschreibung: „Pb C kaut *ängstlich* an seinem Füller.“ – Oder: „Pb C sitzt mit *fiebrigem* Kopf über seiner Testaufgabe.“ – „Er neigt *schielend* den Kopf zum Nachbarn.“

Während in den beiden ersten Fällen verhältnismäßig eindeutige Qualifikationen gegeben werden, dürfte im dritten Beispielfall zwischen verschiedenen Beobachtern bzw. Protokollauswertern ein Konsens nicht mehr so leicht herbeizuführen sein. insofern hier bereits erst Schlußfolgerungen auf die den Handlungen zugrunde liegenden Motive und Ursachen in die Beschreibung einfließen. Sind Schweißausbrüche, Gesichtsrötung und Augentränen Anzeichen für einen ängstlichen oder einen kranken Schüler? Deutet das Kopfnicken die Absicht zu mogeln an? Eine Entscheidung hierüber sollte möglichst nicht schon in die Beschreibungsphase (vor)verlegt werden; sie ist Anliegen der eigentlichen Interpretationsphase.

3. Das *adjektivische* Niveau. Hier setzt nun mehr oder weniger deutlich die Beurteilung ein, wie nachstehende Beispielsätze veranschaulichen: „Pb A ist eine furchtsame (ängstliche) Schülerin.“ – „Pb B ist pedantisch (nörglerisch).“ – „Pb C ist stark aufgeregt (erhältet versus unsicher versus leistungsschwach).“ Was eigentlich geschehen ist, erfährt man hier erst gar nicht mehr. Viele Protokolle über Verhaltensbeobachtungen sind voll von solchen „Vor“-Urteilen und somit als deskriptive Urteilsgrundlage unbrauchbar, sie sind Ausfluß mehr oder weniger subjektiver – verfrühter – Deuteversuche.

4. Das *substantivische* Niveau. Auf dieser Ebene münden Verhaltens-„Beschreibungen“ vollends in bloße Klassifikationen, wie sie für Typologien kennzeichnend sind. Protokollsätze wie „Bei Pb A handelt es sich um eine Phobikerin“ oder „Pb B ist ein Pedant (Querulant)“ oder „Pb C ist ein nervöser Charakter. Er zeigt deutliche Züge einer Leistungsinsuffizienz“ usw. vermitteln kaum noch konkrete oder genauere Vorstellungen über das Gemeinte. Demnach sind auch Validitätskontrollen hier nicht mehr möglich. Die Gefahr, daß es zu Fehlurteilen oder (psychodiagnostisch gesehen) zu Fehldiagnosen bzw. -prognosen kommt, ist zu hoch, um dieses Vorgehen als legitimes Verfahren – selbst bei routinierten Beobachtern – gelten zu lassen. Vor allem aber sollte sich der in der Verhaltensbeobachtung noch ungeübte Testleiter stets der aufgezeigten Gefahren bewußt sein und nicht falschen Leitbildern (aus der Praxis) verfallen.

Die Beschreibung des Beobachteten muß also auf den beiden ersten Verbalisierungsstufen erfolgen. Eine Verhaltensdeskription auf dem 3. oder gar 4. Niveau verbietet sich nach dem Postulat phänomengetreuer Beobachtung bzw. sorgfältiger (nachprüfbarer) Beschreibung des Beobachteten. Die eigentliche *Interpretation*, d.h. die psychologische, nosologische oder wie immer geartete Deutung der Beschreibungsbefunde, schließt sich in praxi vielfach mehr oder weniger nahtlos an die Deskriptionsphase an. In dieser letzten Phase erfolgt dann auch die Integration der Beobachtungsdaten in die Testbefunde (vgl. Gutachten).

Speziell bei Intelligenz- und Schuleignungsuntersuchungen wird die Verhaltensbeobachtung die Testergebnisse im Einzelfall *korrigieren* oder *ergänzen*. So kann eine geringere, deutlich unter der Erwartung liegende, *Testleistung* fallweise als unterdurchschnittliche Begabung (oder Overachievement) versus – mit Hilfe der Verhaltensbeobachtung – (angemessener) als Ergebnis momentaner Indisponiertheit (Unpäßlichkeit, Nervosität, Prüfungsangst u.ä.) interpretiert werden. Andererseits muß der (Test-)Diagnostiker bestrebt sein, die Urteilsbasis (Dateninformationen) möglichst breit anzu-



legen. Dabei werden einmal verschiedene *Tests* qua standardisierte Verhaltensstichproben (vgl. Testdefinition) und zum andern Verhaltens*beobachtungen* (in „freier“ oder „gebundener“ Form) die notwendige Aspektfülle garantieren und gleichzeitig die Zuverlässigkeit der Aussagebefunde durch Bestätigung versus fehlende Bestätigung einzelner Symptome oder Indikatoren erhöhen. Auch in der Diagnostik gilt: *Ein* Verfahren ist *kein* Verfahren! Die Auswahl der Untersuchungsverfahren hängt freilich zuallererst vom Untersuchungsziel – und darüberhinaus möglicherweise vom vorhandenen Methodenangebot – ab.

Weitere Fragen des Untersuchungsdesigns werden wir im empirischen Teil dieses Werkes ansprechen. An Hand dort beschriebener Untersuchungsansätze mag auch das bisher Skizzierte veranschaulicht werden. Ergänzend zu unseren Ausführungen seien hier die Veröffentlichungen von Hasemann (1964), Graumann (1966), Klinghammer (1969) Thoma (1970), Donat (1970) und – für Sonderpädagogen – Bleidick (1970) empfohlen.

## b) Fragebogen (Erhebungs- und Beurteilungsbogen)

Als Instrument der Datenerhebung kommt dem Schüler- bzw. Lehrerfragebogen (seltener auch Elternfragebogen) in der testdiagnostischen Untersuchung nach wie vor große Bedeutung zu. Hierbei ist zwischen dem Fragebogen (FB) als *Erhebungsbogen* im engeren Sinne (z.B. zur Erfassung relevanter Milieudaten, biographischer oder Personaldaten) und dem Fragebogen als *Beurteilungsbogen* im diagnostischen Sinne (etwa zur Veri- oder Falsifikation bestimmter Testbefunde) zu unterscheiden, wiewohl nicht selten beide Formen in einer Liste vereinigt sind. Freilich birgt ein solches Vorgehen gewisse Gefahren in sich. Während der Fragebogen nach dem ersten Modus (FB als Mittel reiner Datenerhebung) ziemlich unproblematisch ist, kann dies vom Fragebogen der zweiten Art (insofern hier nicht selten Beschreibungs- und Beurteilungsphasen miteinander verquickt werden bzw. faktisch mehr oder weniger die Deskriptionsphase fehlt) nicht ohne weiteres behauptet werden. Grundsätzlich sollte man deshalb auch bei Verwendung des Fragebogens als Beurteilungsinstrument nicht auf die Stufe der Verhaltens*beschreibung* (vgl. S. 84f.) verzichten. Das FB-Inventar entspricht hier der gebundenen Form einer Verhaltensbeobachtung, wobei die einschlägigen Beschreibungs- und/oder Beurteilungsdimensionen in eine sog. *Kategorienliste* zusammengefaßt sind. Bei der Zusammenstellung solcher Kategorienlisten ist darauf zu achten, daß erstens die in der Liste vorgegebenen Beobachtungs- bzw. Beurteilungsdimensionen alle wahrscheinlichen Verhaltensmöglichkeiten repräsentieren und zweitens die solcherart gewonnenen Beschreibungen resp. Beurteilungsbefunde hinreichend reliabel sind. Um die Reliabilität (Zuverlässigkeit) der FB-Informationen zu erhöhen, erfolgt die Deskription bzw. Einschätzung oder Beurteilung des Verhaltens von Probanden häufig mit Hilfe mehrstufiger Skalen (z.B. immer – oft – manchmal – selten – nie). M.a.W.: Die Beschreibung oder Einschätzung des Schülerverhaltens an Hand vorgegebener Fragebogenitems sollte möglichst einfach gemacht werden, um so das Agreement der Beurteiler und damit die FB-Zuverlässigkeit optimal zu halten.

Für jede größere schuldiagnostische Erhebung oder auch öftere Einzel-(Intelligenztest-) Untersuchung sollte der Untersuchungs- oder Testleiter (Lehrer) einen Kurzfragebogen entwerfen. Dieser wird je nach Zweck und Ziel der diagnostischen Erhebung inhaltlich und formal modifiziert werden müssen. Bei Begabungsuntersuchungen im Rahmen der baden-württembergischen Bildungsberatung (Schullaufbahnberatung) hat sich etwa folgende FB-Anlage – als Fragebogenmuster der gemischten Form – bewährt. Auf einem einzigen Blatt finden sich hier für jeden Schüler getrennt aufgeführt die wichtigsten Informationen einschließlich der für die Beratung ausschlaggebenden Testbefunde<sup>47</sup>; siehe Tab. 1.

<sup>47</sup> Die im Beispielfall aufgeführten Testverfahren (PSB von W. Horn, AzN 4+ von E. Hylla & B. Kraak, CFT von R.B. Cattell, WST 5/6 bzw. WST 7/8 von H. Anger et al.) werden im empirischen Teil dieses Buches (vgl. Kap. II) beschrieben.

Muster eines Erhebungs- und Beurteilungsbogens (FB-Kurzfassung) für die Bildungsberatung, hier Schullaufbahnberatung im 4. Grundschuljahr – vgl. Aurin et al. 1968, Heller 1968 u. 1972 c

87

Häufig werden freilich diese Standardangaben nicht genügen. So verwandten Aurin (1966) und Heller (1969, 1970) im Rahmen der Ermittlung sog. Begabtenreserven wesentlich detailliertere Fragebogen. Dabei sollten u.a. die Lehrer bezüglich der getesteten Schüler Angaben machen über: 1. *Erscheinungsbild und körperliche Verfassung*, z.B. körperliche Entwicklung, durchgemachte Erkrankungen, Lateralität, Sprachfehler, Hör- und Sehbeeinträchtigungen, Mehrfachbehinderungen u.dgl.m. 2. *Begabung und Leistungsfähigkeit*, z.B. Begabungsgrad, Begabungsrichtung, geistige Wendigkeit, Anregbarkeit bzw. Spontaneität des Denkens, Zielgerichtetheit bzw. Art des Denkverlaufs (bei Problemlösungen), Originalität des Denkens, Kritikfähigkeit, Aufmerksamkeit und Konzentration, Arbeitsweise (Arbeitshaltung), Ausdauer, Belastbarkeit, Leistungsmotivation oder auch Bildungswilligkeit; 3. *Sozialverhalten*, z.B. Kontakt zu Mitschülern, Einordnungsbereitschaft, Selbständigkeit u.ä. 4. *Schülerpersönlichkeit*, z.B. Selbstkritik, Selbstwertgefühl, Temperament, Emotionalität (Affektivität), Willenssteuerung u. dgl.m. Ferner wurden differenzierte Informationen eingeholt über die 5. *Schülerbiographie*, 6. den *schulischen Werdegang*, 7. die besonderen (auch außerschulischen) *Interessen*, die 8. *Bildungsempfehlungen* des Klassenlehrers sowie 9. *Familienverhältnisse* im einzelnen. Ein solch umfangreicher (4- bis 5seitiger) Lehrerfragebogen wird jedoch nur in Ausnahmefällen (besonderen Problemfällen) resp. für spezifische (z.B. wissenschaftliche) Untersuchungszwecke angebracht – und für die betr. Lehrer zumutbar – sein. In der Alltagspraxis genügen meistens 1- bis 2seitige FB-Kurzfassungen der in Tab. 1 dargestellten Art, vielleicht in dieser oder jener Abwandlung. Stärkere Modifikationen werden in der Regel bei *sonderpädagogischen* Fragestellungen notwendig, wie der in Tab. 2 auszugsweise wiedergegebene „Anamnese- und Protokollbogen für sehgeschädigte Kinder“ von H. Horn (1968) beispielhaft demonstriert. Aus Platzgründen kann der umfangreiche Fragebogen hier nur sehr gerafft dargestellt werden. Analoge FB-Konzepte müßten für die (intelligenz)diagnostische Untersuchung Hörgeschädigter, Sprachbehinderter oder Lernbehinderter (z.B. Bleidick 1970) verwendet werden.

#### Tabelle 2.

Auszugsweise Wiedergabe des „Anamnese- und Protokollbogens für sehgeschädigte Kinder“ nach H. Horn (1968, S. 96-100)

*Personaldaten*, z.B. Name, Geschlecht, Geburtsdatum, Tag der Einschulung usw.

*Verlauf der Schwangerschaft und Geburt*, z.B. häufiges morgendl. Erbrechen der Mutter während der Schwangerschaft, Frühgeburt, Normalgeburt, erschwerte Geburt infolge ..., Geburtsgewicht usw.

*Eintritt der Sehschädigung*: bereits bei der Geburt vorhanden / 0;0-0;5 / 0;6-0;11 / 1;0-1;5 / 1;6-1;11 / 2;0-2;5 / 2;6;-2;11 / 3;0-3;11 / 4;0-4;11 / ... / unbekannt

*Augenbefund rechts*, z.B. Anophthalmus, Chorioretinitis, retrolentale Fibroplasie usw. (ophthalmologische Angaben)

*Augenbefund links* (ophthalmologische Angaben)

*Ätiologische Angaben*, z.B. erbbedingt, Rh-Faktor, Toxoplasmose usw.

*Erbbiologische Daten*: gleiches Augenleiden bei Vater / gleiches Augenleiden bei Mutter / gleiches Augenleiden bei einem oder mehreren Geschwister(n) / gleiches Augenleiden bei Großvater väterlicherseits usw.

*Visusangaben*: Besseres Auge (Fern- bzw. Nahvisus), schlechteres Auge (Fern- bzw. Nahvisus), Korrekturen, Bemerkungen über evt. Gesichtsfeldeinschränkungen, Skotome usw.

*Zusatzbehinderungen des Kindes*, z.B. Imbezillität oder Idiotie, Lernbehinderung (Debilität), Hörschädigung, Dysmelien usw.

*Auffälligkeiten des Kindes*, z.B. Übergewicht, Appetitstörungen, Erbrechen, Bettnässen, Einkoten, gestörter Schlafrhythmus, Asthma bronchiale, Epilepsie usw.

*Angaben über Familienverhältnisse und Erziehungssituation*, z.B. Intaktheit, sozio-ökonomische Lage, Schulbildung der Eltern, Geschwisterkonstellation, Erziehungsstil, vor-schulische Betreuung, Sonderbeschulung u.dgl.m.

### c) Psychodiagnostisches Gespräch (Exploration und Anamnese)

Bei Begabungs- bzw. Intelligenztestuntersuchungen der verschiedensten Zweckbestimmung steht am Ende meist ein kürzer oder länger gehaltenes psychodiagnostisches Gespräch. Im Gegensatz zur *testpsychologischen* Standardsituation, wo Untersuchungsleiter und Proband nur in mittelbaren Kontrakt treten, wird das *Psychodiagnostische Gespräch* als (unmittelbarer) *interpersonaler Prozeß* – und somit vielfach als wünschenswerte Ergänzung zur eigentlichen Testuntersuchung – verstanden. Nach Schraml (1964) fallen hierunter die *Exploration* als Tätigkeitsbegriff (Explorieren = gezielte Befragung als Form des Informationsgesprächs) und die *Anamnese* als Gegenstandsbegriff (Anamnesis = Erinnerung; ursprünglich in der klinischen Psychologie bzw. Medizin Bezeichnung für „Krankengeschichte“, d.h. Ausfragen oder Eruiieren der Symptome und begleitenden Umstände (einer Krankheit), allgemeiner das Ingesamt der Mitteilungen des Probanden oder anderer Informanten, z.B. Kindeseltern, Angehörige, Klassenlehrer, Vertrauensarzt). Der Terminus „Psychodiagnostisches Gespräch“ fungiert somit nicht nur als Oberbegriff, diese Bezeichnung ist auch neutraler als die stark klinisch gefärbten Inhalte der Begriffe „Exploration“ und „Anamnese“, die sich freilich fast – wie es scheint – unverzichtbar in die Literatur eingebürgert haben.

Das psychodiagnostische Gespräch als Methode in der (Intelligenz-)Testuntersuchung ist vor allem durch zwei Ziele gekennzeichnet: erstens als *Mittel zur Informationssammlung* (information getting interview) und zweitens als *Instrument der Persönlichkeits- und Eignungsdiagnostik*, d.h. im Hinblick auf schulische Belange vor allem der Begabungsdiagnostik und Schulleistungsprognostik (assessment and evaluative interview). Die dritte Zweckbestimmung, nämlich die *therapeutische* oder *pädagogische* Beeinflussung im Sinne einer Persönlichkeits- oder Verhaltensänderung (educational, consulting and therapeutic interview) gehört nicht mehr in den engeren Bereich diagnostischer Fragestellungen. Das Gespräch als therapeutisches oder erzieherisches Mittel spielt zudem in der Intelligenzdiagnostik allenfalls gelegentlich eine Rolle und kann deshalb im folgenden übergangen werden.

Hinsichtlich des Freiheitsgrades unterscheidet Schraml drei Formen psychodiagnostischer Gesprächsführung: a) das (voll)standardisierte Interview, b) das halbstandardisierte Interview, c) das freie Gespräch. Unter ersteren Gesichtspunkten wäre das Psychodiagnostische Gespräch praktisch den Fragebogen zuzuordnen; zumindest bestehen zwischen einem Anamnese- und Protokollbogen der in Tab. 2 aufgeführten Art und dem voll- resp. halbstandardisierten Interview keine prinzipiellen Unterschiede.

Das Psychodiagnostische Gespräch als Teil einer (intelligenz)testdiagnostischen Untersuchung beginnt gewöhnlich mit der (totalen oder partiellen) *biographischen* Anamnese. Dabei stehen Aussagen des Probanden und seiner Angehörigen resp. wichtigsten Kontaktpersonen im Mittelpunkt des Interesses. Aber auch *Dokumente* (Lebenslauf, Zeugnisse u.ä.) sowie gegebenenfalls *persönlich-intime Produktionen* (z.B. Tagebücher, Briefe, Aufsätze) des Probanden werden in der psychodiagnostischen Untersuchung als (mögliche) Informationsquellen verwertet. Nicht selten – in der *sonderpädagogischen Diagnostik* fast durchweg – spielen in der Intelligenzdiagnostik *medizinische* bzw. *nosologische* Daten eine gewichtige Rolle. Und schließlich werden die auf solche Weise gesammelten Informationen häufig noch durch eigene oder fremde, d.h. durch dritte Personen angestellte, Verhaltensbeobachtungen (meist im Sinne von Gelegenheitsbeobachtungen) abgerundet, wie ja überhaupt in praxi die Informationsmedien „Verhaltensbeobachtung“, „Fragebogen“ und „Psychodiagnostisches Gespräch“ mehr oder minder *integrative* Bestandteile der testdiagnostischen Erhebung darstellen.

Nachstehend bringen wir drei Fallbeispiele, die das bisher (unter Pkt. III, 2) Erörterte konkretisieren sollen. Die Beschreibungen sind Teile beliebig ausgewählter Schülerbeurteilungen, wie sie kürzlich im Rahmen eines Gutachtenseminars am Sonderpädagogischen Institut der Pädagogischen Hochschule Heidelberg erarbeitet wurden. Obwohl es sich bei den Demonstrationsfällen um sehgeschädigte Probanden handelt, können sie als mehr oder

weniger typische Beispiele (sonder)schulpädagogischer Problemdarstellung im Rahmen intelligenzdiagnostischer Entscheidungssituationen angesehen werden. Auch hier geht es ja meist um die Analyse von Schul- und Lernleistungsschwierigkeiten, Verhaltensstörungen usw. oder die Klärung der Schuleignung, also um Fragen der angemessenen Bildungsförderung i.w.S. Sonderpädagogische Fallbeispiele wurden hier vorab unter dem Gesichtspunkt anamnestischer Fragestellung – und der Notwendigkeit ihrer ausgedehnten Berücksichtigung im Rahmen der klinischen und sonderpädagogischen Diagnostik – gewählt. In der intelligenztestdiagnostischen Praxis innerhalb der Regelschule wird man sich vielfach mit einem stark reduzierten Psychodiagnostischen Gespräch oder gar mit der Beantwortung einer Fragebogenkurzfassung (siehe oben) zufrieden geben können<sup>48</sup>.

### *Fallbeispiel A (Carola)*

#### 1. *Anlaß* der Untersuchung:

Carola besucht die 5. Klasse (im 5. Schuljahr) einer westdeutschen Blindenschule. Ihre schulischen Leistungen waren bisher befriedigend. Seit Beginn des 5. Schuljahrs ist ein auffallender Leistungsrückgang festzustellen. Die psychodiagnostische (Intelligenztest-) Untersuchung soll die Ursachen für den Leistungsabfall aufdecken.

#### 2. *Anamnese* (Schülerbogen, ehemalige Klassenlehrerin, jetziger Klassenlehrer, Heimerzieherinnen):

Carola wurde am 14.1.1960 in B. geboren; beidseitiger Anophthalmus.

Vater: Hanspeter H., Busschaffner; Mutter: Lore H., Krankenschwester.

Die Ehe wurde 1 1/2 Jahre nach Carolas Geburt geschieden. Carola wuchs bei ihrer Großmutter väterlicherseits auf. Der Vater hat das Sorgerecht für Carola. Die ehemalige Klassenlehrerin vermutet, daß die Mutter Trinkerin ist. Der Vater hat 1964 wieder geheiratet. Carola blieb aber im Haushalt der Großmutter. Seit ihrer Einschulung wohnt Carola im Internat der Blindenschule und wird ab und zu am Wochenende von ihrer Großmutter abgeholt.

Carola hatte in den ersten vier Schuljahren ein sehr herzliches Verhältnis zu ihrer Klassenlehrerin, die jetzt in eine andere Stadt gezogen ist. In diesen Jahren hatte Carola befriedigende Schulleistungen, in Rechnen sogar gute Leistungen. Auf allen Zeugnissen ist vermerkt, daß Carola sehr zurückhaltend sei und sich reger am Unterricht beteiligen müsse.

Im Herbst 1971 bekam Carola einen neuen Klassenlehrer und verschiedene neue Lehrer für die Fächer Englisch, Erdkunde, Geschichte und Maschinenschreiben. Carolas Schulleistungen sanken seitdem stetig, in Englisch und Maschinenschreiben auf „mangelhaft“. Der neue Klassenlehrer hat den Eindruck, daß Carola die ihr gestellten Aufgaben gar nicht begreift. Sie ist im gesamten Unterricht völlig passiv. Er meint, daß Carola lernbehindert sei und die guten Zeugnissensuren der Vorjahre auf das Wohlwollen und Mitleid der (weggezogenen) Klassenlehrerin zurückzuführen seien.

#### 3. *Beobachtung* bei der Testdurchführung:

Carola läßt sich gern testen. Anscheinend freut sie sich, daß sich jemand mit ihr beschäftigt. Sie sitzt gelöst, mit verschränkten Armen auf ihrem Stuhl. Wenn sie eine Frage nicht gleich beantworten kann, faßt sie sich mit beiden Händen an die Augenlider.

---

<sup>48</sup> Auf die Wiedergabe der intelligenztestdiagnostischen Befunde wurde an dieser Stelle verzichtet. Dieser vielfach wichtigste Teil eines Schülergutachtens wird später noch ausführlich zu besprechen sein (vgl. Gutachtentechnik). Hier interessierten zunächst die Möglichkeiten der Komplettierung testdiagnostischer *Erhebungsdaten* im Hinblick auf die Schülerbeurteilung. Die aufgeführten Protokollauszüge wurden von den Studierenden H. & R. Reich, M. Sappler, C. Steppich und I. Zierer angefertigt; für die freundliche Überlassung der Unterlagen sei ihnen hier gedankt.

### *Fallbeispiel B (Günter)*

#### 1. *Anlaß* der Untersuchung:

Günter wurde im Sept. 1970 in die 1. Klasse für Lernbehinderte an einer westdeutschen Blindenschule aufgenommen. Anfangs der leistungsschwächste Schüler zählt er jetzt zu den besten seiner Klasse. Wegen Erkrankung des Klassenlehrers wurde er wiederholt in der 1. Volksschulklasse mitunterrichtet. Dabei zeigte sich, daß er in seinen Leistungen kaum unter dem Durchschnitt dieser Klasse, d.h. nicht erheblich unter Grundschulniveau, liegt. Es stellte sich deshalb die Frage, ob der Schüler überhaupt lernbehindert sei bzw. ob man Günter zurück in die Volksschule versetzen könne.

#### 2. *Anamnese* (Befragung der Fürsorgestelle der Schule, der Heimerzieherinnen, der Klassenlehrer in der Lernbehinderten- und Volksschule):

##### a) Lebenslauf / Entwicklung:

Günter wurde als lediges Kind am 3.10.1962 in M. geboren. Als Günter noch zu Hause war, arbeitete die Mutter stundenweise als Zugehfrau. Die Großmutter beaufsichtigte das Kind.

Nach Aussagen der Fürsorgestelle wollte die Mutter, um das Blindengeld für sich behalten zu können, Günter nicht zur Schule schicken, sondern als bildungsunfähig bezeichnet wissen.

Den Eindruck mangelnder Bildungsfähigkeit machte Günter auch, als er im Sept. 1970 mit 7;11 J. eingeschult werden sollte. Aufgrund der Retardierungserscheinungen wurde er vorläufig in die Lernbehindertenkasse aufgenommen. Günter näßte nachts und tags ein, nahm nur flüssige Nahrung zu sich. Verhielt sich völlig passiv, bewegte sich kaum, faßte nichts an, sprach nicht und verhielt sich jeder Ansprache gegenüber abweisend. Kontakt mit Kindern und Erwachsenen lehnte er ab.

Mittlerweise (etwa ab Dezember 1970) hat sich sein Verhalten in jeder Beziehung geändert. Er wirkt aufgeschlossen und hat nun ein gutes Verhältnis zu Mitschülern, Lehrern und Erziehern.

Nach Aussagen des jetzigen Klassenlehrers (der Grundschule) liegen seine Leistungen etwas unter dem Durchschnitt der Klasse, was vor allem auf die schleppende Arbeitsweise Günters zurückgeführt wird.

##### b) Augenärztliches Gutachten:

Seit Geburt Opticusatrophie und Amaurose an beiden Augen.

##### c) Zeugnisbemerkung des 1. Halbjahres:

Günters Verhalten ist sehr eigenwillig, manchmal aggressiv. Es hat sich jedoch im Lauf des Einschulungsprozesses gebessert. Er konnte allmählich zu geordneter Mitarbeit geführt werden und zeigt beachtliche Lernerfolge. Besonders zu vermerken ist sein sprachliches Ausdrucksvermögen.

### 3. *Verhaltensbeobachtung*:

#### a) Beobachtung bei der Testdurchführung:

Günter verstand oft die Fragestellung nicht. Für ein Versagen, wenn auch nur vermeintlich, entschuldigte er sich: „Alles weiß ich nicht.“ Er fing ab und zu an zu singen, wollte mit dem Testmaterial spielen, wollte manches behalten, gab es aber sofort wieder her, wenn er dazu aufgefordert wurde. Er ahmte die Geräusche nach, die er hörte.

#### b) Ausdrucksbewegungen:

Günter zeigt keine der sog. Blindeneigenheiten, weder Augenbohren noch Schaukeln. Er hat beim Gehen eine normale Haltung. Die Augen hat er immer weit aufgerissen

### *Fallbeispiel C (Klaus)*

#### 1. *Anlaß* der Untersuchung:

Klaus besucht die 8. Klasse einer westdeutschen Sehgeschädigtenschule. Sein Verhalten stellte Lehrer und Erzieher vor die Frage: Kann der Schüler in der Schule verbleiben oder muß er in einer Nervenheilanstalt untergebracht werden?

## 2. Anamnese:

### a) Lebenslauf:

Klaus wurde am 12.5.1955 in A. als Sohn des staatenlosen Ausländers Sandor H. (Ungar), Hilfsarbeiter, und der Deutschen Margarete H., Putzfrau, geboren. Die Eltern sind inzwischen geschieden. Der Aufenthalt des Vaters nach seiner Entlassung aus der Strafanstalt in S. ist unbekannt. Die Frau ist inzwischen wieder verheiratet und hat noch zwei gesunde Kinder. In die Schule trat Klaus am 9.9.1962 ein.

### b) Augenärztliches Gutachten:

Beidseitiger Mikrophthalmus, Augenzittern; Ursachen: Angeborener Hirnschaden; Grad der Sehschädigung: Klaus nimmt nur Lichtschein wahr.

### c) Ärztlicher Befund (des Hausarztes): Epilepsie

### d) Auszüge aus dem Gutachten des Max-Planck-Instituts für Psychiatrie (kinderpsychiatrische Abteilung):

Der jugendliche Patient wurde wegen Verhaltensauffälligkeiten vorgestellt. Seit 1/2 Jahr ist der Junge bei neuen Situationen extrem ängstlich, beim Zahnarzt hat er nahezu einen Ausnahmezustand bekommen. Vor den Kindern hat er keine Angst.

Parallel dazu sind die Schulleistungen, die vorher im Durchschnittsbereich lagen, etwas abgefallen. Er ist aber seit 1/2 Jahr in einer anderen Gruppe, in der auch größere Kinder sind, die ihn teilweise sehr "trätzen". Bei den häuslichen Besuchen an jedem Wochenende gibt es mit den Geschwistern, denen sich Klaus unterlegen fühlt, gewisse Eifersüchteleien.

Nach den uns zur Verfügung stehenden Unterlagen wurde Klaus 2 Monate vor dem errechneten Termin mit einer rechtsseitigen Amaurose und einem linksseitigen Katarakt geboren. Mit zwei Jahren lernte er sprechen und mit 2 1/2 Jahren laufen.

Nach dem augenärztlichen Gutachten von 1967 besteht rechts Amaurose mit Phtisis bulbi und Mikrophthalmus. Auf dem linken Auge wird noch etwas Licht gesehen (unter 1/100).

**Befunde:** Der schwache, untergewichtige Junge zeigt eine gewisse Dyskranie (pathologische Schädelkonfiguration): schmalen Schädel (die Röntgenaufnahme zeigte nur 14,5 cm Durchmesser), niedrige Stirn, Überbiß.

Internneurologisch waren keine pathologischen Befunde zu erheben. Das EEG war bis auf eine nicht ganz stabile Grundtätigkeit unauffällig. Es bestanden keine großen Asymmetrien.

Wir führten den HAWIK-Verbalteil durch; der Pb erreichte einen IQ von 92 Punkten, also ein leicht unterdurchschnittliches Ergebnis.

Bei der Testdurchführung war Klaus extrem ängstlich, auf energisches Zureden hin jedoch kooperativ.

Zusammenfassend können wir sagen, daß es sich bei K. um eine angeborene Amaurose unklarer Genese handelt. Wir glauben, daß es sich bei den Verhaltensstörungen um reaktive Störungen auf die Versetzung in eine andere Gruppe handelt. Außerdem sind einige Verhaltensweisen durch die beginnende Pubertät erklärbar.

Mit Lehrern und Erziehern wurden einige Erziehungsprobleme besprochen.

### e) Schulleistungen:

In der 7. Klasse (1969/1970) betrugen die Zensuren in den Fächern Kurzschrift, Lesen und Singen 2, in Betragen, Fleiß, Religion und Naturlehre 3, in Deutsch, Erdkunde, Geschichte, Sozialkunde, Naturkunde, Werken und Englisch 4, in Rechnen und Turnen 5 sowie in Maschinenschreiben 6. Die Zeugniszensuren in der 8. Klasse (1970/71) waren fast durchweg, in den Fächern Geschichte und Maschinenschreiben sogar um zwei Notennstufen, besser. – Die 1. (Grundschul-)Klasse hatte K. wiederholt.

### f) Eintragungen im Schülerbogen (vom 22.7.1970):

Fleiß und Betragen des Schülers waren zufriedenstellend.

Klaus litt in letzter Zeit zunehmend unter teils heftigen epileptischen Anfällen, die eine Spezialbehandlung nötig machten. Er ist allgemein leicht erregbar und leidet teilweise an Lern-, Gedächtnis- und Bewußtseinsstörungen als unmittelbare Folge eines Anfalls. Seine kindlich-naive, verantwortungsfremde Haltung gegenüber der Schule und ihren Anforderungen kommt erschwerend hinzu. Besondere Erwähnung verdient sein erstaunliches Zahlengedächtnis.

### g) Eintragungen im Schülerbogen (vom 15.2.1971):

Klaus leidet zunehmend an starken epileptischen Anfällen. Er schaltet manchmal völlig ab und kann sich stundenlang nicht recht konzentrieren. Seine leichte Reizbarkeit und heftigen Wutausbrüche machen ihn oft zur Zielscheibe seiner Mitschüler. Trotz seiner schlechten körperlichen Konstitution kam Klaus doch zu erfreulichen Erfolgen, zumal er das Wort Hausaufgabe anscheinend noch nie gehört hat. Er interessiert sich sehr für Literatur und manuelle Tätigkeit. Sein enormes Merkvermögen übersteigt zeitweise alle Erwartungen.

### h) Aussagen des Erziehers:

Klaus stört ständig das Gruppenleben; er provoziert die anderen und ist unordentlich. Sein Verhalten ist unkontrolliert und kann gefährlich werden („Er ging einmal auf mich (Erzieher) mit dem Messer los!“). Andererseits ist Klaus überaus hilfsbereit; im Grunde sucht er Anerkennung.

K. hat jetzt täglich einen Anfall: 3 - 4 Min. mit Schreien und Weinen, danach ist er ohnmächtig. Nach den Ferien sind die Anfälle regelmäßig schwerer, verbunden mit einer nachfolgenden völligen Ermattung für 3 - 5 Stunden. Offenbar nimmt K. zuhause seine Tabletten nicht regelmäßig.

### i) Aussagen der Mutter:

Klaus schlägt seine Schwester und drosselte sie sogar schon. Sein kleiner Bruder (1;6 J.) ist durch ihn gefährdet: K. schüttelt ihn und geht überhaupt sehr grob mit ihm um. K. wird deshalb nicht mehr allein mit seinem jüngeren Bruder gelassen. Trotzdem sind Stiefvater und Mutter um Klaus besorgt. Das derzeitige Verhalten ist ihrer Meinung nach eine vorübergehende Erscheinung.

### j) Vorkommnisse in der Schule bzw. im Schulinternat:

Am 4.3.1971 stieß K. einem Mitschüler bei der Tischmahlzeit die Gabel in die Wange; die Verletzung machte ärztliche Behandlung notwendig.

Am 26.3.1971 warf K. sein Messer über die Köpfe der Mitschüler im Speisesaal. Auch früher hat K. schon mit einem Messer geworfen.

Am 30.3.1971 zertrümmerte K. einen Teller auf dem Kopf eines Mitschülers, nachdem er bereits Tassen und Teller vom Tisch gefegt hatte.

Am 17.4.1971 stieß K. einen gehbehinderten Mitschüler so stark an die Wand, daß dieser stürzte und sich verletzte.

Ferner wird berichtet, daß K. bereits Einrichtungsgegenstände (z.B. Schreibmaschinen) oder Eigentumsgegenstände von Mitschülern (Tonbandgerät) im Gesamtwert von rd. 1000 DM zerstört habe. Auch würde er wiederholt seine Mitschüler treten und ohrfeigen.

### 3. Verhaltensbeobachtung:

#### a) Beobachtungen während der Testdurchführung:

Klaus zeigt sich sehr wissensdurstig und mitteilbar. Überhaupt wollte er lieber mit dem Tester plaudern als sich den geforderten Testaufgaben zuwenden. Während der eigentlichen Testdurchführung schweifte er ständig ab, so daß der Zeitaufwand (für die Testung) außergewöhnlich groß war. Andererseits erwies er sich als gutmütig und zeigte sich fortwährend behilflich; so wollte er ständig Bücher, die Tasche und ähnliche Gegenstände (des Testleiters) tragen.

#### b) Ausdrucksbewegungen:

Klaus zeigt typische Blindismen: Augenbohren, Körperwiegen u.ä. Sein Gang ist aber sicher, sein Orientierungsvermögen recht gut. Häufig hat er beim Gehen die Arme angewinkelt, wobei die Hände auffällig schlaff herabhängen.

#### c) Sonstige Beobachtungen:

Klaus zeigt ein starkes Verlangen nach Kontakt. So hat er die Tendenz, die Erwachsenen, mit denen er zusammenkommt, ständig zu berühren.

In den vielfach gerühmten Vorteilen – Unmittelbarkeit des Kontakts und Natürlichkeit der Situation – des Psychodiagnostischen Gesprächs liegen zugleich seine Gefahrenquellen, die hier nur kurz skizziert werden sollen. Die größte Gefahr liegt wohl in der *Subjektivität* der Aussagen resp. der via Exploration erhaltenen Informationen: „Wir erfahren, wie der Pb sich



selbst, sein Verhalten, seine Beziehung zur Umwelt sieht bzw. erlebt; objektivierbare Daten (Schulleistung, Anzahl der Freunde ...) werden zwar auch erhoben, aber letztlich werden auch sie subjektiv verarbeitet mitgeteilt" (Sommer 1971, S. 195 f.). Neben subjektiven Wahrnehmungen, Erlebnissen usw. können aber auch Gedächtnisfehler oder „Erinnerungslücken“ – alle individuellen Wahrnehmungs- und Gedächtnisleistungen sind gewissen Selektionstendenzen unterworfen – den Wert der mitgeteilten Informationen infragestellen, von möglichen Täuschungsabsichten (bewußter oder unbewußter Art) seitens des Probanden oder bestimmter Informanten einmal abgesehen.

Auf der anderen Seite können beim Untersuchungsleiter (Testleiter in der Rolle des Beobachters und/oder Explorators) selbst gewisse *Informationsverzerrungen* auftreten. „Aufgrund bestimmter Einstellungen und Erwartungen können Äußerungen des Pb ungenau wahrgenommen werden (Informationen werden falsch verstanden, „überhört“ oder „hineingehört“). Weiterhin können das Verhalten des Untersuchungsleiters allgemein oder seine Fragetechnik im besonderen zu Fehlerquellen werden. Ein distanziertes, mürrisches, aber auch ein zu freundliches Auftreten kann den Pb in seiner Äußerungsbereitschaft beeinflussen. Durch die Formulierung der Fragen selbst (z.B. Suggestivfragen, die die Antwort des Pb in eine bestimmte Richtung drängen), aber auch durch intensives oder nachlässiges Fragen bei bestimmten Bereichen sowie durch unterschiedliche Sachkenntnis kann der Untersuchungsleiter ebenfalls das Ergebnis des Gesprächs verfälschen" (Sommer, loc. cit.).

Schließlich spielen *Sympathie* versus *Antipathie* im interpersonalen Prozeß, d.h. hier in der unmittelbaren Begegnung von Test- oder Untersuchungsleiter einerseits und Proband bzw. Informant andererseits, nicht selten eine Rolle im Hinblick auf das Ergebnis der Befragung, auch der anamnestischen Erhebung (siehe oben). Ferner wissen wir heute, angeregt durch die psycholinguistische Forschung, daß selbst das Medium Sprache – auch bei adäquater Beschreibung des Verhaltens in formaler Hinsicht – Gefahrenquellen in sich birgt; wir meinen die *Gefahr sprachlicher Mißverständnisse*, die durch unterschiedliche (z.B. sozio-kulturell bedingte) Verhaltenscodes gegeben ist (vgl. die früheren Ausführungen zum "restringierten" und "elaborierten" Sprachcode auf Seite 38ff.).

Zur Verminderung der aufgezeigten Fehlerquellen wurden verschiedene Lösungsvorschläge unterbreitet: Protokollierung des Psychodiagnostischen Gesprächs durch dritte Personen, Aufzeichnung durch Tonband oder Video-Rekorder, wiederholte Gespräche in bestimmten zeitlichen Abständen und ähnliche Objektivierungshilfen, die jedoch allenfalls eine graduelle Minderung der Gefahren bedeuten. Hinzu kommt der spürbare Mangel hinreichend abgesicherter *Auswertungskriterien* explorativ erhobener resp. anamnestischer Dateninformationen. Diese Kritik darf aber nicht so verstanden werden, daß hinfort auf explorative oder anamnestische Datenerhebungen verzichtet werden sollte. Vielmehr gilt es m.E. darauf zu achten, die aufgezeigten Fehlerquellen auf ein Minimum zu reduzieren, um psychodiagnostische Gesprächsinformationen als wirkliche Hilfen testpsychologischer Untersuchungen auszuweisen. Diese Forderung gilt auch im Hinblick auf den Einsatz von Exploration und Anamnese im Bereich schulischer Intelligenzdiagnostik, wenngleich diese Methoden dort nie die Bedeutung erlangen wie vollstandardisierte Intelligenz- und Leistungstests.

### 3. Das pädagogische Gutachten (Integration der Testbefunde und Dateninformationen – Beratungshinweise)

Der schwierigste Teil psychologisch-diagnostischer Untersuchungen ist zweifellos die Phase der *Interpretation* der Test- bzw. Datenbefunde und ihre Integration zum sog. *Gutachten*, d.h. die Erarbeitung eines abschließenden Urteils. Im Rahmen intelligenzdiagnostischer Untersuchungen in der Schule impliziert dies die

Forderung nach einer objektiven, reliablen und validen *Schülerbeurteilung* im Hinblick auf die jeweils gegebene Problemsituation (z.B. Bedingungsanalyse der Schulleistung, Erhellung der kognitiven Voraussetzungen bestimmter Lernleistungen, Ermittlung der Schuleignung u. dgl. m.). Beurteilungen dieser Art stellen dann die Grundlage dar für notwendige Entscheidungen, wie sie vom Lehrer in der Schule oder auch vom Schüler selbst bzw. den Eltern außerhalb der Schule gefordert werden. Ferner liefern solche Beurteilungen die notwendigen Informationen für eine gleichermaßen wirklichkeitsangemessene und zukunftsorientierte Bildungsplanung. Im Zentrum der Bemühungen steht demnach die *Bildungsberatung* als übergreifende Aufgabe *individualpsychologischer* Beratung (z.B. Beratung individueller Lern- und Verhaltensschwierigkeiten, Schullaufbahnberatung, individuelle Beratung von Schülern, Eltern und Lehrern) und *informativer* Beratung (im Rahmen von Schulmodellversuchen, Bereitstellung notwendiger Dateninformationen für die Bildungsplanung usw.). *Bildungsberatung ist somit ein Strukturelement des modernen Bildungswesens*, wie der Deutsche Bildungsrat in einer seiner jüngsten Empfehlungen formulierte (vgl. Strukturplan für das Bildungswesen vom 13. Februar 1970, S. 91 ff.).

Beratungen dieser Art sind ohne geeignete *Methoden* nicht denkbar, wie im gleichen Bericht mehrfach unterstrichen wird, z.B.: „Eine individuelle Schullaufbahnberatung gründet sich auf Gespräche mit Lernenden, Eltern und Lehrern und auf Tests, die Kenntnisse und Fähigkeiten, Eignungen und Einstellungen, Interessen und Neigungen ermitteln“ (S. 92). Der schulische Einsatz von *Intelligenztests* i.w.S., d.h. Tests, die die gesamte Kern- und Sekundärintelligenz umfassen (s.S. 11f.), bildet somit eine wesentliche Grundlage für eine effiziente Beratung im Bildungswesen. Darauf sind wir bereits früher (s.S. 81ff.) ausführlich eingegangen. Hier sollte nur noch einmal der Zusammenhang von Beurteilung und Beratung herausgestellt werden. M.a.W.: Psychologisch-diagnostische Untersuchungen in der Schule sind nie Selbstzweck; sie stehen im Kontext je anhängiger Problemsituationen, zu deren Lösung sie beitragen sollen. Damit aber ist eine wichtige Kennfunktion des pädagogischen Gutachtens schon benannt: Jedes Gutachten ist durch seine *Zweckbestimmung* gekennzeichnet (vgl. u.a. Heiß 1964, Hartmann 1970).

Unter dem Kriterium der Zweckbestimmung unterscheidet Heiß drei Grundformen des Gutachtens: erstens das (umfassende) *Persönlichkeitsgutachten*, d.h. „das Gutachten als Darstellung und ‚Bild‘ der Persönlichkeit“, zweitens das *Problemgutachten*, d.h. „das stellungnehmende und urteilende Gutachten“, drittens „das beratende Gutachten“ (S. 976). Der Beratungsaspekt ist jedoch direkt oder indirekt immer relevant, wie wir oben betonten, so daß sich eigentlich nur zwei Grundtypen, das Persönlichkeitsgutachten und das Problemgutachten, herauschälen lassen. Beim *pädagogischen* Gutachten steht indessen das Problemgutachten im Vordergrund, so daß wir uns hier mit der Erstellung eines sog. Persönlichkeitsbildes – das zudem in der psychologischen Diagnostik heute nicht unproblematisch ist – nicht zu befassen brauchen. Allgemein wird deshalb das *Gutachten* definiert als *die zusammenfassende Darstellung der Untersuchungsbefunde*, wobei man sich auf ein oder – in der Regel – mehrere Verfahren stützt (vgl. Heiß 1964)

Während Veröffentlichungen über Tests bzw. psychodiagnostische Probleme in kaum mehr überschaubarer Fülle vorliegen und noch ständig zunehmen, mutet die Literatur zum Thema „Gutachten“ – umfangmäßig und inhaltlich – recht bescheiden an. Vor allem fehlt eine umfassende und für die Praxis verbindliche Theorie des Gutachtens. Wer die einschlägige Literatur (z.B. Hammond & Allan 1953, Klopfer 1960, Huber 1961, Heiß 1964, Schubenz 1967, Thomae 1967, Hartmann 1970) studiert, findet zwar eine Fülle von Anregungen und mehr oder weniger praktikablen Empfehlungen, „die aber vielfach recht unverbindlich bleiben und in der Regel über technische Ratschläge nicht hinausgehen. Selten finden sich grundsätzlichere Überlegungen, die die Problematik des Gutachtens und seine Bedeutung im Rahmen der psychologischen Diagnostik klar herausstellen und Ansatzpunkte für eine entsprechende Grundlagenforschung bieten. Dabei wäre für die Psychodiagnostik nichts wichtiger als eine Systematik des Gutachtens und nichts so irrig wie die Annahme, man könne sich zu diesem Zweck auf die Erfahrung stützen und auf eine methodisch kontrollierte Bedingungsanalyse verzichten“ (Hartmann 1970, S. 71). Dies gilt uneingeschränkt oder sogar in verstärktem Maße für das *pädagogische* Gutachten.

Trotzdem (oder gerade deshalb) soll hier die Methodik des Gutachtens wenigstens nach den mir besonders relevant erscheinenden Aspekten pädagogisch-psychologischer Beurteilung versus Beratung angesprochen und an Hand einiger Fallbeispiele wiederum konkretisiert werden.

Am Ende jeder Testuntersuchung bzw. Datenerhebung steht die Aufgabe, die Fülle der Informationen teilweise recht heterogener Provenienz (z.B. aus Tests, Verhaltensbeobachtung, Fragebogen, Anamnese oder Exploration) so zu systematisieren, daß daraus eine echte Entscheidungshilfe für die anstehende Problemlösung wird. Zweck des Gutachtens ist ja letzten Endes eine Entscheidung, z.B. über die Schullaufbahn, die erforderlichen pädagogischen Maßnahmen nach erfolgter Schulleistungsanalyse, bestimmte didaktische Hilfen bei Lernschwierigkeiten u.ä. Dabei wird eine solche Entscheidung häufig vorgegebenen oder gewählten Maßstäben folgen, etwa der Altersgruppen- oder Klassennorm, dem gesteckten Lernziel oder allgemeinen Leistungsanforderungen eines bestimmten Bildungsweges (Schuleignungskriterien). Diese *Maßstaborientierung* ist kennzeichnend für das Prüfungs- und Urteilsgutachten, d.h. praktisch für alle intelligenzdiagnostisch relevanten Urteile in der Schule. Unter dem Gesichtspunkt der *Beratung* wird dann freilich das Individuum selbst mehr oder weniger zum Maßstab; hierzu ist eine verstehende und einfühlende Haltung seitens des (Bildungs-)Beraters notwendig, wobei eine gründliche Situationsanalyse, etwa in der Prognostik die sorgfältige Abwägung des wahrscheinlichen – künftigen – situativen Umfeldes, zumeist unerlässlich wird. Gerade aber die Kenntnis situativer Variabler fehlt dem Gutachter resp. Berater nicht selten, was die Hauptschwierigkeit längerfristig gültiger Prognoseurteile (z.B. am Ende der Grundschule die Vorhersage des späteren Bildungserfolgs auf Gymnasium oder Realschule) ohne weiteres aufzeigt. Andererseits müssen auch hier wieder optimale Gütekriterien der verwendeten Verfahren (Erhebungs- und Verarbeitungstechniken) als die besten Voraussetzungen eines treffsicheren diagnostischen oder prognostischen Urteils im Sinne vorhin skizzierter gutachtlicher Stellungnahme und Beratung angesehen werden.

Integrative Bestandteile der Gutachtenerstellung sind die Arbeitsphasen der Test- bzw. *Datenauswertung*, der Befund*interpretation* und der Gutachten*formulierung*. Entsprechend bestimmen sich die Konstruktionsprinzipien: „1. Analyse und weitgehende *Formalisierung* von Datenauswertung, diagnostischer Urteilsbildung und Gutachtentechnik. 2. *Interdependenzanalyse* von Untersuchungsplanung, Auswertungstechnik, Urteilsprozessen und Gutachtentechnik. 3. *Operationalisierung* des psychodiagnostischen Vokabulars. 4. Analyse und theoretische Fundierung des *Bezugsrahmens* von Gutachten (diagnostische Erschließung von Persönlichkeits- und Verhaltensmodellen, Situationsanalyse, Auftraggebertypologie u.a.)“ (Hartmann 1970, S. 73 f.).

Bei unserer Erörterung der psychodiagnostischen Grundstruktur wurde schon der Stichprobencharakter der bzw. jeder Testuntersuchung betont. Analog ist die Grundsituation des Gutachtens ausgezeichnet durch den Akt des Deutens und Schlußfolgerns, d.h. hier wird am Ende der psychodiagnostischen Untersuchung immer von einem Teilverhalten auf ein Ganzverhalten geschlossen, also von einer bestimmten Verhaltensstichprobe (im Test, in der Beobachtung usw.) auf die Grundgesamtheit des Pb-Verhaltens generalisiert. So ist etwa bei der Intelligenztestdiagnose weniger das aktuelle Intelligenzniveau als vielmehr die relativ überdauernde intellektuelle Leistungsfähigkeit des Schülers, bei einer Schuleignungsprognose der tatsächliche spätere Bildungserfolg in dieser oder jener Schulform gefragt. Um nun solche Aussagen zu ermöglichen, müssen auf Gutachtenniveau die verschiedenen Testbefunde, Fragebogen-, Beobachtungs- und/oder Anamnesedaten in einem konstruktiven Akt zu einem Ganzen verbunden, d.h. systematisch geordnet werden. Dabei stellen sich nach Heiß (1964, S. 987 ff.) drei Aufgaben für den Gutachter:

1. *Bewertung der Einzelbefunde*. „Der Gutachter muß sich darüber im klaren sein, welches Gewicht er dem jeweiligen Befund gibt. Handelt es sich nur um einen einzigen Befund, so ist zu entscheiden, wie dieser Befund innerhalb des Gesamtbildes steht, welches im Gutachten erscheint. Handelt es sich aber um verschiedene Befunde, so müssen diese Befunde nach Aussagegewicht und Aussagewert miteinander verglichen werden“ (loc. cit.). In der Regel wird man davon ausgehen können, daß dominante Züge, Eigenschaften, Fähigkeiten u.ä. einer Person mehr oder weniger markant in den Testbefunden resp. Dateninformationen aus Fragebogen, Verhaltensbeobachtung, Anamnese und Exploration hervortreten. So gibt es Befunde, die

a) überdurchschnittlich (stark) versus b) unterdurchschnittlich (schwach) versus c) durchschnittlich (unauffällig) ausgeprägt erscheinen. Schon ein Vergleich von a) und b) gestattet oft sehr schnell Aufschlüsse über eine bestimmte Strukturierung, etwa Stärken und Schwächen einer Verhaltensweise. „So tritt beispielsweise im Intelligenztest ein auffällig hoher oder niedriger Wert zutage. Alle diagnostischen Verfahren kennzeichnen ihre abartigen und aus der Reihe fallenden Befunde, welche zeigen, daß ein Funktionsbereich stärker hervortritt, sei es im positiven oder störenden Sinne. Selbst in der Exploration werden uns auffällige Verhaltensweisen genannt, wir bemerken bestimmte Ausfallserscheinungen und seien es nur die, daß übliche Verhaltensweisen nicht zu beobachten sind. Macht man einmal den Versuch, nach abgeschlossener Untersuchung lediglich die auffälligen und hervortretenden Befunde den weniger deutlichen, blassen und durchschnittlichen Befunden gegenüberzustellen, so zeichnet sich oft schon eine bestimmte Struktur ab. Es gibt einerseits die Fälle, in denen wir gehäuft auffällige Befunde haben, andererseits wiederum die ganz blassen Erscheinungen, bei denen im Grunde alles innerhalb des Durchschnittlichen und Üblichen bleibt, es gibt weiterhin die Fälle, in denen nur *ein* Befund mit besonderem Gewicht hervortritt“ (loc. cit.). Bei der abschließenden Beurteilung (Gutachtenerstellung) wird man allerdings nicht bei einer solchen Rohgewichtung der Befunde stehen bleiben können. Auf der nächsten Stufe der Befundintegration erfolgt deshalb die syndromatische Organisation.

2. *Die Ordnung der Befunde zu Syndromen.* Hier wird danach gefragt, welche Befunde sich vereinigen lassen, d.h. sich gegenseitig stützen, und welche Befunde kontrastieren. Gleichsinnige Befunde (Symptome) verstärken natürlich ihr Gewicht. „So kann beispielsweise ein hoher Intelligenzbefund, der sich im messenden Verfahren (Intelligenztest; d. Verf.) ergibt, mit dem Intelligenzbefund des Rorschachtests übereinstimmen und weiterhin durch andere Befunde... unterstützt werden. Wenn in dieser Weise eindeutig von verschiedenen Seiten her eine überdurchschnittliche Funktion der Intelligenz bestätigt wird, können wir mit Sicherheit auf diesen Gesamtbefund bauen“ (loc. cit.). Doch gibt es auch sog. Kontrastsyndrome, d.h. Befunde, die sich widersprechen. Sie können eine gewisse Ungleichgewichtigkeit der betreffenden Funktion andeuten bzw. – positiv – auf Beweglichkeit der Person versus – negativ – auf Unbeständigkeit des Verhaltens hinsichtlich einer bestimmten Leistungsdimension hinweisen. Entsprechend sind die Ursachen für solche Befunddivergenzen in Methodendifferenzen – die verwendeten Intelligenztests messen beispielsweise (trotz gleicher Etikettierung) nicht dieselben Fähigkeiten – oder in einer außergewöhnlichen Persönlichkeitsdynamik bzw. in Intelligenz-Leistungsdiscrepanzen (z.B. Underachievement versus Overachievement) u.dgl.m. zu suchen. Während also gleichsinnige Syndrome geschlossene und mehr oder minder reibungslos funktionierende Verhaltensstrukturen indizieren, deuten Kontrastsyndrome sehr oft Spannungen im Persönlichkeits- bzw. Leistungsgefüge an. Schließlich gilt es

3. „*Grundwert*“ und „*Stellenwert*“ der Befunde gegeneinander abzuwägen. „Sowohl die Bewertung der Befunde nach Gewicht wie die Kombination der Befunde, die sich aus verschiedenen Verfahren ergeben, haben eins gemeinsam: sie deuten den zunächst für sich stehenden Befund in seinem Zusammenhang mit den anderen Befunden. Sie geben ihm, verglichen mit den anderen Befunden ein größeres oder geringeres Gewicht oder sie zeigen, wie ein Befund im Zusammenhang mit andern sich abschwächt oder verstärkt. So oder so korrigieren wir also auf diese Weise den ursprünglichen Befund, welchen wir erhalten haben. Der *Grundwert*, welchen wir diesem ursprünglichen Befund zuschreiben können, wird durch einen *Stellenwert* ersetzt... Der *Stellenwert* wird sichtbar, sobald wir den Grundwert im Zusammenhang der anderen, in der Persönlichkeit wirksamen Verhaltensweisen betrachten. Wir nennen ihn deshalb Stellenwert, weil wir mit seiner Bestimmung bereits ein Gesamtgefüge der Person voraussetzen, in welchem nun die ermittelte Verhaltensstruktur an einer bestimmten Stelle des Gefüges steht“ (loc. cit.). M.a.W.: Der Grundwert des ursprünglichen (Einzel-) Befundes avanciert in der Gesamtbewertung nach Gewicht und syndromatischer Ordnung zum Stellenwert, erweist sich in seiner spezifischen Bedeutung im übergeordneten Kontext. Heißt unterscheidet hier zwischen einem „dynamischen“ und einem „statischen“ Stellenwert. Der Stellenwert eines (Test-)Befundes kann sich je nach Variation der äußeren (situativen) und/oder inneren (z.B. motivationalen) Verhaltensbedingungen ändern, was insbesondere bei Leistungs- bzw. Verhaltensprognosen – wie bereits angedeutet – zum

Problem wird. Im allgemeinen ist man in der Diagnostik bzw. Prognostik darauf aus, das durchschnittliche Verhalten unter durchschnittlichen Bedingungen zu bestimmen bzw. vorherzusagen. Diese Faustregel gilt zumindest im Hinblick auf die traditionelle, hier geschilderte, Konstruktion eines Gutachtens. Bei speziellen Fragestellungen, etwa der Schul-eignungsdiagnostik, gewinnen in allerjüngster Zeit noch andere Modelle zunehmende Bedeutung: Wir meinen die probabilistischen Ansätze und – damit eng verbunden – die automatischen Klassifikationskonzepte der Eignungs- und Leistungsdifferenzierung im Bildungswesen (vgl. Allinger und Heller 1972). Aus didaktischen Gründen werden wir jedoch erst später (vgl. Kap. III, 1 u. 2a im empirischen Teil dieses Werkes) hierauf zu sprechen kommen. Nur soviel sei hier bereits angemerkt, daß dort das Abwägen von Grund- und Stellenwert einzelner Befunde in einen objektivierbaren Operationalisierungsprozeß einmündet und somit ein Maximum an Zuverlässigkeit bislang unbekannter Größenordnung erreicht.

Nachstehend aufgeführte Fallbeurteilungen dienen nicht nur der Veranschaulichung, sie sollen zugleich das bisher Gesagte nach der einen oder anderen Richtung hin verdeutlichen. Die Beispiele stellen freilich keine „Ideal“-Gutachten dar; sie repräsentieren wiederum Fallbeschreibungen aus dem früher zitierten Gutachtenseminar an der Pädagogischen Hochschule Heidelberg <sup>49</sup>. Etwaige vom Leser empfundene Nachteile thematischer Art – es handelt sich bei der Probandenauswahl erneut um *sehgeschädigte* Schüler – werden m.E. durch einigermaßen repräsentative Schulprobleme resp. ausführliche Darstellungen im Anamnese- und Beobachtungsprotokoll mehr als wettgemacht. Soweit die erwähnten Testmethoden dem Leser unbekannt sind, sei zur Information auf das sehr detaillierte Methodenverzeichnis auf Seite 113ff. verwiesen. Um dem noch ungeübten Diagnostiker die Gutachten-erstellung transparenter zu machen, wurde in den beiden ersten Fallbeispielen ein Zwischenschritt der Urteilsfindung mitaufgezeichnet: die syndromatische Konstruktion vor dem eigentlichen Abschlußurteil. Die in dieser Systematik sonst nicht übliche Aufführung der Urteilsbelege im Gutachten soll vor allem die Gefahr möglicher Überinterpretationen versus der Vernachlässigung wichtiger Befunde bannen helfen. In den letzten beiden Gutachten wurde auf die Dokumentation der in dieser oder jener Form notwendigen Kontrolle der Urteilsfindung – vorab aus Platzgründen – verzichtet. Dafür werden wir im empirischen Teil eine Reihe weiterer einschlägiger Demonstrationsfälle bieten, allerdings ohne Wiedergabe der Beobachtungs- und Explorationsdaten.

### *Fallbeispiel D (Gabriele)*

#### *1. Anlaß der Untersuchung:*

Gabriele besucht z.Z. der Untersuchung die 3. Klasse einer westdeutschen Sehbehindertenschule und soll an eine Lernbehindertenschule überwiesen werden. Die intelligenzdiagnostische Untersuchung soll Aufschlüsse über die adäquate Beschulung vermitteln, also die Frage der *Lernbehinderung* klären.

#### *2. Anamnese:*

##### *a) Lebenslauf:*

G. wurde am 3.11.1960 in M. als uneheliches Kind geboren. Die Mutter (geb.: 1923), lt. angegebenem Beruf Putzerin, ist fast immer unbekannten Aufenthalts und kümmert sich nicht um das Kind. Der Vater ist nicht bekannt (Angaben des Stadtjugendamtes, das die Amtsvormundschaft hat).

Seit 1962 lebt G. bei Pflegeeltern in B. Der Pflegevater ist von Beruf Schlosser und arbeitet in der Fabrik. Die Pflegemutter ist als Hausfrau tätig. Mit der (leiblichen) zwölfjährigen Tochter der Pflegeeltern versteht sich G. sehr gut. Überhaupt ist G. sehr gerne bei ihren Pflegeeltern, die sie als ihre richtigen Eltern betrachtet (Pb-Aussage).

---

<sup>49</sup>Für die teilweise gerafft wiedergegebenen Gutachtenfälle zeichnen die Studierenden M. Hengstler, M. Schuber, H. Schuster und C. Wiemar, denen hierfür der Dank ausgesprochen sei, freundlicherweise verantwortlich.

Der Pflegemutter ist lediglich bekannt, daß G. 1963 am rechten Auge und 1967 am linken Auge staroperiert wurde. 1967 wurde G. eingeschult. Nach Aussage der Volksschullehrerin machte G. einen gepflegten Eindruck und war immer sauber gekleidet. Im Frühjahr 1970 verschlechterten sich Gabrielles Schulleistungen – vermutlich aufgrund ihrer Sehbehinderung – so sehr, daß die Pb auf Antrag des örtlichen Schulamts in die Sehbehindertenschule eingewiesen wurde.

b) Augenärztliches Gutachten:

Beide Augen staroperiert, Linsen bei beiden Augen operativ entfernt. In der Iris kein operativer Einschnitt.

Sehschärfe mit Glas (Starglas): R = 5/35; L = 5/35.

c) Auszüge aus dem amtsärztlichen Gutachten vom Februar 1970:

Körperlicher Zustand: Altersmäßig entwickeltes Kind, Größe 129 cm, Gewicht 23 kg, kein krankhafter Organbefund. Haltungsschwäche, hoher Gaumen mit Zahnfehlstellung. – Zustand nach Staroperation: Linsenlosigkeit bds.; trotz Starbrille Sehvermögen unter 1/10

Geistig-seelischer Zustand: Normale geistige Entwicklung, normale Begabung. Gutachten der Behörde: Es handelt sich um eine nicht nur vorübergehend hochgradig sehschwache Person. Dauernd erwerbsunfähig, auch im erwachsenen Alter, wenn keine Hilfe erfolgt. Gesamtminderung 100 v.H. Überweisung in eine Sehschwachenschule erforderlich, spätestens mit Beginn des Schj. 1971/72.

d) Schulleistungen:

Schuljahr 1967/68: Leistungen im Schreiben gut, in Lesen und Rechnen zwischen ausreichend und mangelhaft.

Schuljahr 1969/70: Leistungen in Schrift, Heimatkunde und Handarbeit ausreichend, in Deutscher Sprache und Rechnen mangelhaft (Abgangszeugnis der Grundschulklasse 3).

Schuljahr 1970/71: Zensuren nach Repetition in der Sehbehindertenschulklasse 3 in Handarbeit befriedigend, in Rechnen, Heimatkunde und Schrift ausreichend, in Deutscher Sprache mangelhaft.

e) Eintragungen im Schülerbeobachtungsbogen:

**Körperliche Entwicklung:** Normale Körpergröße, schlechte Haltung. Sprachfehler: G. stößt mit der Zunge an. G. leidet an beiden Augen am grauen Star, sie trägt eine sehr starke Brille; kann aber nur lesen, wenn sie das Buch 3 cm nah an die Augen hält. Körperlich ist G. nicht sehr wendig. Manuell – im Zeichnen und Handarbeiten – ist G. aufgrund ihres Augenleidens wenig geschickt.

**Umwelt:** Gabriele lebt in einer Pflegefamilie, die sich sehr um sie sorgt. Die zwei Jahre ältere Pflegeschwester betreut sie sehr liebevoll. Die Familienverhältnisse in der Pflegefamilie sind geordnet. Das Kind kommt immer sauber und gut gekleidet in die Schule.

**Fähigkeiten und Neigungen:** Die ausreichenden bis mangelhaften Leistungen in der Schule beruhen vor allem auf der Kurzsichtigkeit der Schülerin, aber auch auf der kaum durchschnittlichen Begabung.

**Arbeitsweise:** G. ist sehr langsam, unkonzentriert, allzuleicht ablenkbar, trotzdem aber bemüht um Sauberkeit und Sorgfalt in der Arbeitsausführung.

Gabi bräuchte jemanden, der hinter ihr steht, sie aufmuntert, neue Erklärungen gibt.

**Soziales Verhalten:** Fügt sich gut in die Klassengemeinschaft ein. Großer Bewegungsdrang. Freundlich, höflich, gutmütig, gesellig.

f) Angaben der jetzigen Lehrerin und Erzieherin im Schulinternat:

In Anlehnung an Bleidick (1970) wurden folgende *Persönlichkeitsdimensionen* bei der Pb eingeschätzt:

Intelligenz: niedrig–sehr niedrig

Schulleistung: niedrig

Entwicklungsstand: einförmig

Auffassung: schwerfällig

Arbeitsweise: stockend/langsam

Gedächtnis: mangelhaft

Konzentration: ablenkbar

Sprache: dürftig–gestört

Antrieb: mäßig aktiv

Steuerung: gespannt

Stimmung: froh

Selbstgefühl: unsicher

Anregbarkeit: sensibel

Gemüt: weich

Angepaßtheit: bemüht

Halt: unsicher

Soziale Einordnung: anlehnungsbedürftig

Soziale und erziehlche Reaktion: fügsam

### 3. Verhaltensbeobachtung während der Testuntersuchung mit dem INS:

Gabriele wurde am 24.3.1971 mit dem INS – Intelligenztest für normalsichtige und sehgeschädigte Kinder und Jugendliche; deutsche Bearbeitung des Williams Intelligence Test for Children with Defective Vision von Prof. Dr.H. Horn, Abteilung für Heilpädagogik der Päd. Hochschule Ruhr in Dortmund – getestet und erreichte einen vorläufigen *Gesamt-IQ* von 80 ( $\pm 5$ ) IQ-Punkten.

G. tut sich schwer bei der verbalen Formulierung der Antworten. Die Fragen müssen ihr sehr oft vorgesprochen werden. Ihre Leistungen im Behalten von Zahlen und Sätzen sind sehr schlecht (G. kann keinen einzigen Satz und nur mit Mühe 4 Zahlen nachsprechen). Bei den Zusatzaufgaben arbeitet sie langsam und unsicher. Bei jeder Aufgabe fragt sie, ob es so richtig sei u.ä.

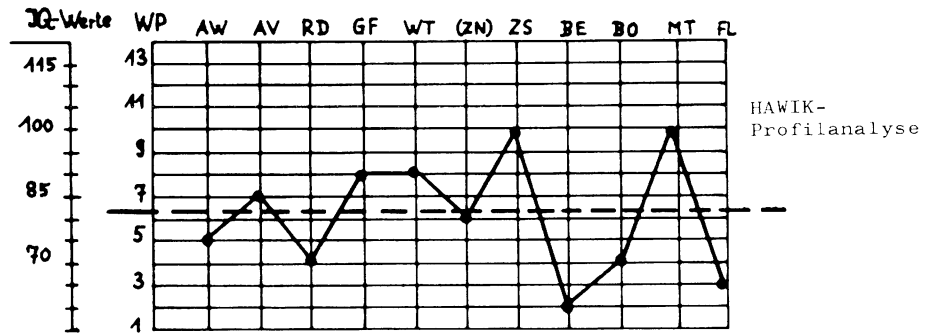
### 4. Ergebnisse der Testuntersuchung mit dem HAWIK am 2.7.1970:

Testalter: 9:7 J.

HAWIK-Verbalteil: Subtests	1. AW (Allgem. Wissen)	5 WP (Wertpunkte)
	2. AV (Allgem. Verständnis)	7 WP
	3. RD (Rechn. Denken)	4 WP
	4. GF (Gemeinsamkeitenfind.)	8 WP
	5. WT (Wortschatztest)	8 WP
	5a. ZN (Zahlennachsprechen)	6 WP
	Verbalpunkte:	38 / 32 WP *
	6. ZS (Zahlensymboltest)	10 WP
	7. BE (Bilderergänzen)	2 WP
	8. BO (Bilderordnen)	4 WP
HAWIK-Handlungsteil:	9. MT (Mosaiktest)	10 WP
	10. FL (Figurenlegen)	3 WP
	Handlungspunkte:	29 WP
	Gesamt-Wertpunkte	61 WP *

Verbal-IQ: 77 \*  
 Handlungs-IQ: 71  
 Gesamt-IQ: 71 \*

Anm.: \* = nach Umrechnung der 6 Verbaltests auf der Basis von 5 Subtests



### 5. Syndromatische Ordnung der Befunde:

Unterdurchschnittliche bzw. niedrige oder sehr niedrige allgemeine Intelligenz; sehr niedriges *aktuelles* Intelligenzniveau.

Prämotorisches Intelligenzniveau (Intelligenzkapazität) vermutlich höher, d.h. im Normbereich oder nur knapp durchschnittlich.

*Intraindividuelle kognitive Fähigkeitsschwerpunkte*

(Stärken): Abstraktionsfähigkeit und Denkkonsequenz; Umstrukturierungsfähigkeit u.ä.

INS: rd 80 IQ

HAWIK-Gesamtergebnis: 71 IQ

Anamn.: Pkt. e, f

MT: 10 WP = 100 IQ

WT: 8 WP = 90 IQ

Anamn.: Pkt. c ?

GF: 8 WP = 90 IQ

MT: 10 WP = 100 IQ

Allgemeine psychomotorische Geschwindigkeit sowie Lernfähigkeit (vorab in bezug auf motorische Handlungen) bzw. visuell-motorische Koordination trotz hochgradiger Sehbehinderung im interindividuellen Durchschnittsbereich, nicht darunter.

Vgl. dagegen Kontrastsyndrom: TU / LU (TU = Testurteil; LU = Lehrerurteil) bzw. LU / LU;

Stützen für TU: Handarbeits- und Schreibnoten (Zensuren) verhältnismäßig gut.

**Intraindividuelle – und interindividuelle – Schwächen**  
intellektueller Leistungsfähigkeit: Rechnerisches Denken, geistige Wendigkeit; Bildungsstandard (von sozio-kulturellen Determinanten stärker abhängig); dagegen müssen die Schwächen im HAWIK-Handlungsteil in erster Linie durch die starke Sehbehinderung der Pb bedingt interpretiert werden: also Visualität und Beobachtung, d.h. visuelles Erkennen bekannter Figuren und Gegenstände, optische Differenzierungsfähigkeit; Klarheit und Schärfe bildlicher Vorstellungen, planende Phantasie u.ä.

Hingegen liegen situatives Verständnis, die Einsicht in Feldzusammenhänge (sog. *praktische Intelligenz* oder praktisches Urteilsvermögen) sowie die Gedächtnisleistungen im HAWIK im intraindividuellen Mittel, wenn auch interindividuell mehr oder weniger deutlich unter dem Durchschnitt (vgl. aber ZS); stützende Hinweise für schlechte *Gedächtnisleistungen* überwiegen jedoch in den Belegquellen.

Auch die Belege für Konzentrationsfähigkeit, Aufmerksamkeit u.ä. Modifikationen des *Arbeits*verhaltens deuten ziemlich gleichsinnig auf entsprechend niedrig ausgeprägte Pb-Merkmale. Einzige Ausnahme ist ZS, wobei aber hier vielleicht stärker der Speedfaktor im Psychomotorischen als Komponente von ZS gewertet werden muß. Ansonsten ist die Pb anstrengungsbereit und um positive Leistungen bemüht.

Insgesamt liegen die *verbalen* Intelligenzleistungen der Pb in der für Sehgeschädigte typischen Weise (vgl. dazu ausführlicher noch Kap. III, 2c im empirischen Teil dieses Buches) *höher* als die Handlungintelligenz, wenngleich auch da im interindividuellen Vergleich deutlich unter dem Altersdurchschnitt.

**Schulleistungsfaktor** im HAWIK: extrem niedrig (spricht für Lernbehinderung). TU und LU (Zensuren) kongruent. (Zur Maßstabsproblematik bzw. Kriterienbestimmung der Lernbehinderung vgl. die Ausführungen in Kap. III, 2 a im empirischen Teil dieses Buches.)

**Sozialität:** insgesamt freundlich und aufgeschlossen, wenn auch mehr passiv, weich, anlehnungsbedürftig, fügsam usw.

Weitere **Auffälligkeiten:** schwaches Selbstwertgefühl, unsicher; braucht Bestätigung, Aufmunterung, päd. Ermutigung.

**Sozio-kulturelles Milieu:** frühkindliche Deprivationen wahrscheinlich; Pb erscheint sozial in die Pflegefamilie voll integriert. Gewisse sprachliche (und kognitive?) Deprivation auch in der Pflegefamilie nicht ausgeschlossen.

**Hirnorganische Störungen:** Organikersyndrom im HAWIK (niedrige Leistungen in den nebenstehenden Subtests) eher negativ, d.h. o.B. – niedrige FL-Leistung ist durch die gravierende Sehbehinderung zu erklären.

WT: 8 WP = 90 IQ

ZS: 10 WP = 100 IQ

MT: 10 WP = 100 IQ

ZS und MT / Anamn.:  
Pkt. e (manuelle Ungeschicklichkeit)

versus Anamn.: Pkt. d /  
Pkt. e

RD: 4 WP = 70 IQ

AW: 5 WP = 75 IQ

Anamn.: Pkt. f

BE: 2 WP = 60 IQ

BO: 4 WP = 70 IQ

FL: 3 WP = 65 IQ

AV: 5 WP = 85 IQ

ZN: 6 WP = 80 IQ

ZS: 10 WP = 100 IQ

Anamn.: Pkt. f

INS + Beob.

RD, ZN, BE

INS + Beob.

Anamn.

ZS

Anamn., Beob.

HAWIK:

V-IQ = 77

H-IQ = 71

Anamn.: Pkt. d, f

INS + Beob.

RD, ZN, AW; ZS ?

Anamn.

Anamn.

Beob.

Anamn.

Beob.

Anamn.: Pkt. a

Anamn., Tests

ZS, FL, MT



## 6. Abschließendes Urteil (*Gutachten*):

Ziel der intelligenzdiagnostischen Untersuchung war hier die Klärung der *Schuleignung*, genauer: die Beantwortung der Frage, ob die Probandin G., die z.Z. die Sehbehindertenschule besucht, in eine *Lernbehindertenschule* überwiesen werden soll. Zu diesem Zweck wurden 2 Intelligenztestbatterien (HAWIK und INS) durchgeführt und zahlreiche Anamnesedaten, ergänzt durch Befragungen und Verhaltensbeobachtung, erhoben. Die Befunde können zusammenfassend folgendermaßen interpretiert werden.

Gabriele läßt sowohl in ihren Intelligenzleistungen als auch in ihrer gesamten Schulleistung verhalten deutliche Entwicklungsrückstände, bezogen auf den Schuljahrgangs- und Klassendurchschnitt Sehender bzw. Sehbehinderter, erkennen. Zwar ist es nach einigen Testergebnissen (z.B. HAWIK: MT, ZS) nicht auszuschließen, daß die Pb bei günstigeren familiären bzw. sozio-kulturellen Bedingungen im Vorschulalter und ohne die vorliegende – im Hinblick auf schulische und außerschulische Erfahrungs- und Lernmöglichkeiten nicht unerhebliche – Sehschädigung noch zu durchschnittlichen Begabungs- und Lernleistungen fähig gewesen wäre. Andererseits stimmen die Befunde in bezug auf die aktuellen kognitiven Lernleistungsvoraussetzungen von G. darin überein, daß die Pb den Lernanforderungen der Regelschule oder auch der Grund- und Hauptschule innerhalb einer Sehgeschädigtenschule jetzt sowie in absehbarer Zukunft kaum genügen wird.

Unter Berücksichtigung der hochgradigen Sehbehinderung, die als primäre Behinderung hier in Erscheinung tritt, sowie im Hinblick auf die pädagogische Notwendigkeit, die Kenntnisse und Lernleistungen der Schülerin, besonders im verbalen und numerischen Bereich, gezielt zu fördern und zu verbessern, wird die weitere Beschulung in einer *Lernbehindertenklasse innerhalb der Sehbehindertenschule* vorgeschlagen. Jedenfalls erschien uns diese Bildungsform unter Berücksichtigung aller Komponenten für Gabriele in der jetzigen Situation als optimale Lösung. Sollte ihre Realisierung aus irgendwelchen, z.B. organisatorischen, Gründen nicht möglich sein, würden wir für eine (weitere) Beschulung innerhalb der Sehbehindertenschule plädieren, wobei allerdings spezielle Stützmaßnahmen unabdingbar wären.

Neben individuell abgestimmten pädagogisch-didaktischen Konzepten und Unterrichtstechniken müßten parallele Maßnahmen zur Stabilisierung der Schülerpersönlichkeit, vorab zur Festigung des Selbstvertrauens, und zur aktiven sozialen Integration – gerade auch bei einer eventuellen Eingewöhnung in eine neue Klassen- und Internatsgruppe – geplant werden. Dabei sollte man sich des offensichtlich positiven Einflusses der Pflegefamilie, die entsprechend zu informieren wäre, helfend vergewissern. Persönlichkeits- und sozialpsychologische Stützmaßnahmen der aufgezeigten Art werden nicht zuletzt auch der Schulleistungstüchtigkeit Gabriele zugute kommen. Sofern eine schulpsychologische Beratungsstelle der Schule, die Gabriele besucht oder besuchen wird, angeschlossen ist, wird diese zweckmäßig die weitere Betreuung der Pb übernehmen. In jedem Falle sollte in regelmäßigen Abständen – etwa halbjährlich – durch Retestungen die weitere Begabungs- und Schulleistungsentwicklung Gabriele beobachtet werden, um die besprochenen Maßnahmen gezielt einsetzen zu können und gegebenenfalls (bei deutlichen Lernfortschritten) die Frage der angemessenen Beschulungsform erneut zu überprüfen.

### *Fallbeispiel E (Inge)*

#### *1. Anlaß der Untersuchung:*

Inge besucht z.Z. der Untersuchung die 6. Klasse einer Sonderschule für Sehbehinderte und Blinde. Inge möchte nach dem 7. Schuljahr die Studienanstalt in Marburg (Gymnasium für blinde bzw. sehgeschädigte Schüler mit Internat) besuchen. Eine Entscheidung des derzeitigen Klassenlehrers bezüglich der Schuleignung Inges wurde noch nicht getroffen. Die von der Pb besuchte Sehbehindertenklasse besteht erst seit einem Jahr. Nach ihrer Umschulung aus der Volksschule in die genannte Sonderschule für Sehgeschädigte besuchte Inge 1 Jahr lang eine gemischte Klasse (mit blinden *und* sehbehinderten Schülern) und wurde dort mehr blinden- als sehbehindertenpädagogisch unterrichtet. Wegen des kurzen Beobachtungszeitraums war es dem Klassenlehrer nicht möglich, ein Urteil über Inges *weiterführende Schuleignung* abzugeben. Dieses Gutachten soll diese Frage klären helfen.

## 2. Anamnese:

### a) Lebenslauf:

I. wurde am 30.9.1959 als erstes Kind der Eheleute A. in L. geboren. Ihre Entwicklung verlief normal. 1965 wurde sie in die Grundschule (Volksschule) in B. eingeschult. 1963 machten sich erste Anzeichen einer Sehbehinderung bemerkbar. 1968 mußte I. sich einer Gehirnoperation (Tumor) unterziehen. Die Sehbehinderung schritt rasch voran. Nach dem Krankenhausaufenthalt besuchte I. wieder die Volksschule. Im Entwicklungsbericht vom 12.7.1969 heißt es: „Sie holte den versäumten Stoff schnell nach. Sie gibt trotz Behinderung nicht auf.“ Die Sehbehinderung wurde jedoch so gravierend, daß I. vom 1.9.1969 an die Sonderschule für Blinde und Sehbehinderte besuchen mußte. Inge hat noch eine jüngere Schwester, die ab Herbst des kommenden Jahres auf das Gymnasium übertreten möchte. Die Geschwister verstehen sich gut untereinander, ebenso mit den Eltern.

### b) Augenärztliches Gutachten (1968):

Operation eines chiasmanahen Tumors. Bisher sehr gute Erfolge, mit Ausnahme der erwarteten und inzwischen eingetretenen zunehmenden Optikusatrophie bds. mit Visusreduktion. Mit weiterer Visusreduktion muß gerechnet werden. Gesichtsfeldeinengung: links mehr als rechts. Mit Wachstum verbliebener Tumorreste oder Entstehen neuer Tumore ist zu rechnen. VR 1/20; VL 1/25.

Nach Angaben weiterer ärztlicher Gutachten ist seit 1969 keine Visusverschlechterung mehr eingetreten. Allerdings ist Inges Sehschwäche im Sinne des BSHG § 24, Abs. 2 als Blindheit einzustufen: „Als blind im Sinne des Gesetzes gilt auch, wer eine so geringe Sehschärfe hat, daß er sich in einer ihm nicht vertrauten Umgebung ohne fremde Hilfe nicht zurechtfinden kann.“

### c) Schulzensuren (Zeugnisnoten im 1. Halbjahr des Schuljahrs 1970/71):

Auf der 6. Klassenstufe (6. Schj.) erhielt Inge folgende Zensuren: Text. Gestalten 1; Religion, Rechtschreiben, Geschichte, Erdkunde, Biologie, Sozialkunde, Mathematik, Physik, Bildner. Erziehung und Werken jeweils 2; Deutsch (Hauptnote), mündl. Ausdruck, schriftl. Ausdruck, Lesen, Schreiben, Englisch und Turnen jeweils 3; Singen 4; Blinden-Kurzschrift und Schwimmen teilgenommen.

## 3. Personbeschreibung und Verhaltensbeobachtung:

Inge ist von normaler, d.h. altersgemäßer Körpergröße, im Verhältnis zu ihrer Größe ist die Pb jedoch übergewichtig. Ihre Bewegungen wirken oft schwerfällig und träge. Manuell ist I. sehr geschickt; sie fertigt gerne Handarbeiten an. Wegen der vorhandenen Optikusatrophie trägt die Pb eine Brille.

In der Schule arbeitet I. ruhig und konzentriert. Bei Lehrern und Schülern ist I. gleichermaßen beliebt. Die Pb verhält sich im Internat genauso wie in der Schule unauffällig. Bei der Testung arbeitete I. konzentriert.

## 4. Ergebnisse der Testuntersuchung mit dem HAWIK am 21.6.1971 und dem WST 5-6 am 23.6.1972:

Testalter: 11;8 J.

HAWIK-Verbalteil:	1. AW (Allgem. Wissen)	14 WP
	2. AV (Allgem. Verständnis)	16 WP
	3. RD (Rechn. Denken)	14 WP
	4. GF (Gemeinsamkeitenfind.)	17 WP
	5. WT (Wortschatztest)	13 WP
	5a. ZN (Zahlennachsprechen)	11 WP
	Verbalpunkte:	85 / 71 WP *
	6. ZS (Zahlensymboltest)	8 WP
	7. BE (Bilderergänzen)	11 WP
	8. BO (Bilderordnen)	2 WP
HAWIK-Handlungsteil:	9. MT (Mosaiktest)	14 WP
	10. FL (Figurenlegen)	8 WP
	Handlungspunkte:	43 WP
	Gesamt-Wertpunkte:	114 WP *

Verbal-IQ: 126 \*  
Handlungs-IQ: 90  
Gesamt-IQ: 111 \*

Anm.: \* = nach Umrechnung der 6 Verbaltests auf der Basis von 5 Subtests

Frankfurter Wortschatztest WST 5-6 (Testdauer hier 60 Min.) in Punkschriftdarbietung – vgl. S. 146f. sowie Kap. III, 2c im empirischen Teil dieses Buches:

Ergebnis: 40 RP (Rohpunkte) = 85. PR (Prozentrang) oder 116 IQ

### 5. Syndromatische Ordnung der Befunde:

Intelligenzniveau überdurchschnittlich gut; die im HAWIK ermittelte Diskrepanz zwischen Verbal- und Handlungs-IQ zuungunsten der Handlungsintelligenz ist in erster Linie durch die starke Sehbehinderung der Pb zu erklären, d.h. der HAWIK-Handlungsteil ist für Sehbehinderte kein adäquates Diagnostikum (vgl. S. 210ff. unten).

HAWIK-Gesamtergebnis:  
111 IQ  
V-IQ = 126  
H-IQ = 90  
WST: 85. PR od. 116 IQ  
MT: 14 WP/120 IQ

#### HAWIK-Profilanalyse:

Insgesamt typisches HAWIK-Profil Sehgeschädigter (vgl. Kap. III, 2c im empirischen Teil).

HAWIK-Profil (Subtests)

#### Leistungsschwerpunkte im HAWIK:

Abstraktlogisches Denken und verbale Begriffsbildung sehr hoch ausgebildet. Damit einhergehend recht guter Bildungsstandard und gute Wissensbreite der Pb. Als sehr gut ist die Fähigkeit der Pb zu bezeichnen, frühere Erfahrungen auszuwerten und zu verbalisieren, d.h. ihre praktische Urteilsfähigkeit und ihr situatives Verständnis (Einsicht in Feldzusammenhänge).

GF: 17 WP/135 IQ  
AW: 14 WP/120 IQ  
WT: 13 WP/115 IQ  
WST: 85. PR/116 IQ

Die ausgeprägten Fähigkeiten der Pb im formallogischen Denken zeigen sich auch bei der Lösung von rechnerischen Problemen (Textrechnungen); neben dem formallogisch-adäquaten Denken sind besonders gut entwickelt die Fähigkeiten zur Abstraktion, das Differenzierungsvermögen und die räumliche Vorstellung.

AV: 16 WP/130 IQ  
RD: 14 WP/120 IQ  
MT: 14 WP/120 IQ

*Intraindividuell schwächer*, aber interindividuell im Durchschnittsbereich gelegen, sind gewisse Arbeitshaltungsfaktoren, z.B. Aufmerksamkeit und Konzentration. Allerdings muß bei der Bewertung der HAWIK-Subtestleistungen in ZS, BE und FL die starke Sehbehinderung der Pb unbedingt berücksichtigt werden (s.o.), so daß die entsprechenden Testergebnisse bei einem anderen – sehbehindertenadäquaten – Diagnostikum zur Messung der genannten Eigenschaften sehr wahrscheinlich besser ausgefallen wären als es hier der Fall ist. M.a.W.: Die Arbeitshaltung und Konzentration der Pb dürfte tatsächlich über den hier (im HAWIK) ermittelten Werten liegen. Dafür spricht auch die Beobachtung des Klassenlehrers und der Testleiterin sowie die Tatsache, daß Inge nach ihrer schweren Krankheit (Tumoroperation) den versäumten Lernstoff sehr rasch nacharbeitete. Die HAWIK-Leistungen in BO sind wegen der hochgradigen Sehbehinderung der Pb nicht interpretierbar.

ZN: 11 WP/105 IQ  
ZS: 8 WP/90 IQ  
BE: 11 WP/105 IQ  
FL: 8 WP/90 IQ

Die Schulleistungszensuren entsprechen in etwa den intellektuellen Lernvoraussetzungen (Begabung der Pb), ausgenommen die sprachl. Schulnoten – dort würde man vielleicht im Hinblick auf die im Intelligenztest ermittelten verbalen Fähigkeiten der Pb etwas bessere Noten erwarten.

Beob.  
Anamn. a)  
Anamn. c)

BO

Anamn. c)

*Oberschulsyndrom*: positiv, d.h. gymnasialgeeignet lt. Testbefund (hohe Werte in nebenstehenden HAWIK-Subtests).

AW: 120 IQ  
RD: 120 IQ  
GF: 135 IQ  
WT: 115 IQ

## 6. Gutachten:

Nach den vorliegenden Untersuchungsbefunden ist Inge ohne Einschränkung als *gymnasialgeeignet* im Sinne der weiterführenden Schulbildung anzusehen. Im Hinblick auf die Sehbehinderung der Probandin wird der Übertritt in die Studienanstalt (Gymnasium) für Sehgeschädigte in Marburg vorgeschlagen.

### Fallbeispiel F (Wolfgang)

#### 1. Anlaß der Untersuchung:

W. besucht z.Z. der Untersuchung die Abschlußklasse der Wirtschaftsschule (Sonderschule für Blinde) und möchte zum Schuljahrende auf die Untersekunda der Marburger Studienanstalt überwechseln. Das Gutachten soll die *Gymnasialeignung* des Pb klären.

#### 2. Anamnese:

##### a) Lebenslauf:

W. wurde am 9.3.1953 geboren, ist z.Z. der Testung 18 Jahre alt; die Mutter ist als Hausmeisterin tätig, der Vater infolge eines Kriegsleidens 1958 gestorben. Der Pb hat eine Zwillingsschwester, die normal sieht.

W. ist groß und von kräftiger Gestalt, gesund. Die Ursache der Geburtsblindheit ist unbekannt (Amaurose mit Lichtschein).

##### b) *Explorationsdaten* (Aussagen der Lehrer, Erzieher u.a. Kontaktpersonen sowie des Probanden selbst):

**Kognitive Funktionen:** Nach Übereinstimmung aller Befragten verfügt W. über eine sehr hohe Intelligenz. Er fasse sehr rasch auf, denke logisch, arbeite konzentriert, habe ein ausgezeichnetes Gedächtnis und sei sprachlich sehr gewandt. Praktische Arbeiten lägen ihm dagegen weniger.

Der Pb selbst sieht seine Schwerpunkte auf sprachlichem Gebiet sowie in den Sachfächern (Geschichte, Politik). Mathematische Zusammenhänge durchschaue er schwer. Er arbeite nur so gründlich als notwendig, sein Gedächtnis leiste ihm häufig gute Dienste.

**Emotionale Funktionen:** Die Befragten bezeichneten W. übereinstimmend als ernst, zurückhaltend, ziemlich verschlossen, stets unauffällig, früher geradezu kontaktscheu (manche Befragten geben dieser Entwicklung der Internatssphäre eine gewisse Mitschuld).

Der Pb selbst bezeichnete als seine Grundstimmung: ernst, besorgt, zurückhaltend, emotional sehr ansprechbar (eines seiner Hauptprobleme bestünde darin, seine Gefühle stets in den Griff zu bekommen); er sei gerne mit Menschen zusammen, suche Kontakt; das Leben sei aber nicht leicht, er sei von einem „gesunden Pessimismus“ erfüllt.

**Praktische Funktionen:** Hier traten größere Unterschiede in der Beurteilung zutage: Ein Informant bezeichnete W. als einen Mitläufer, der sich an andere anähne und sich teiben lasse, ein anderer nannte ihn einen „stillen Mitläufer“. Ein Pädagoge sprach von ihm als einem sehr kritischen, idealistischen jungen Mann, der „extrem links“ eingestellt sei, von seinem einmal bezogenen Standpunkt nicht mehr abweiche, sondern sich übertrieben festlege. Wieder ein anderer Informant wollte dieses „extrem links“ nicht gelten lassen; er sprach von „sozialkritischer Einstellung“ und erwähnte auch, daß W. bei seinen Mitschülern beliebt sei, sie durch seine rasche Auffassungsgabe und die flinke Arbeitsweise häufig überflüge.

Der Pb selbst verabscheut nach eigenen Worten bloße Mitläufer. Er habe ein „gesundes Selbstvertrauen“, er wisse, was er zu tun habe, selbständig wolle er seinen Weg im Leben gehen. Mitleid brauche er nicht. Auf autoritäres Verhalten reagiere er allergisch. Er müsse viel lernen. Nach seiner Berufsausbildung als Soziologe wolle er mithelfen, unser Gesellschaftssystem hier in der BRD zu verändern. Jedoch lehne er Gewalt ab, er sei für friedliche Entwicklung. Ihm schwebte eine Art „demokratischer Sozialismus“ vor, der allerdings erst in Ansätzen (z.B. in Chile) zu erkennen sei. Er freue sich sehr auf Marburg, wo er hoffe, mehr Diskussionspartner im Hinblick auf seine Ideen zu finden als hier in X.

### 3. *Verhaltensbeobachtung* (während des Unterrichts, der Testuntersuchung, der Exploration):

Es entstand der Eindruck, daß es sich bei W. um einen sehr intelligenten Pb handelt, der zunächst sehr zurückhaltend und verschlossen wirkt, der aber durchaus fähig ist, aus sich herauszugehen (vgl. Exploration) und mit Begeisterung und Überzeugungskraft seine Ideen darzustellen vermag.

**Zeugnisnoten:** Religionslehre, Geschichte, Wirtschaftsrechnen und Betriebswirtschaft jeweils 1; Deutsch, Englisch, Gemeinschaftskunde, Physik, Chemie usw. jeweils 2, Mathematik, Maschinenschreiben, Briefwechsel und Turnen jeweils 3.

### 4. *Ergebnisse der Testuntersuchung* mit dem HAWIE und dem HISAB (für Blinde adaptierte Form des HAWIE-Handlungsteils) im Herbst 1970:

Testalter: 17;8 J.

HAWIE-Verbalteil: Subtests	1. AW (Allgem. Wissen)	13 WP
	2. AV (Allgem. Verständnis)	12 WP
	3. ZN (Zahlennachsprechen)	16 WP
	4. RD (Rechnerisches Denken)	11 WP
	5. GF (Gemeinsamkeitenfinden)	13 WP
	Verbalpunkte:	65 WP
	5a. WT (Wortschatztest)	13 WP

Verbal-IQ: 128

Anstelle des HAWIE-Handlungsteils wurde hier der HISAB-Test (Haptic Intelligence Scale for Adult Blind von Shurrager) durchgeführt. Dabei erzielte der Proband folgende Resultate.

HISAB-Subtests	1. ZS (Zahlensymboltest)	13 WP
	2. MT (Mosaiktest)	12 WP
	3. FL (Figurenlegen)	8 WP
	4. BE (Bilderergänzen)	8 WP
	5. PB (Pattern Board)	10 WP
	6. BA (Bead Arithmetic)	16 WP
	Handlungspunkte:	67 WP

Handlungs-IQ: 106

### 5. *Gutachten:*

Gegen die Eignung des Pb W. für die weiterführende Gymnasialbildung an der Studienanstalt in Marburg liegen keinerlei Bedenken vor. Sowohl die bisherigen Schulleistungen als auch die Testergebnisse im HAWIE und HISAB, besonders in den Denkfunktionen bzw. sprachlichen Dimensionen, lassen neben guten oder sogar ausgezeichneten Lernleistungsvoraussetzungen i.e.S. (Gedächtnis, Leistungsmotivation, Bildungsinteresse) einen erfolgreichen Studienabschluß erwarten.

### *Fallbeispiel G (Sabine)*

#### 1. *Anlaß* der Untersuchung:

Sabine besucht z.Z. der Untersuchung die 1. Grundschulklasse (Repetition im 2. Schuljahr) und hat Schwierigkeiten im Lesen und Rechtschreiben. Es besteht der Verdacht auf *Legasthenie*. Die Testuntersuchung soll hierüber Aufschlüsse vermitteln bzw. die Frage klären, ob S. lernbehindert ist.

#### 2. *Anamnese:*

##### a) Biographisches:

Sabine wurde am 30.10.1961 in S. geboren und hat noch zwei Geschwister – Bruder von 19 J. und Schwester von 14 J. –, die die Hauptschule mit sehr gutem Erfolg absolviert haben sollen. S. war eine Zangengeburt (Angabe der Mutter) und ist kurzsichtig (progressive Myopie, bds. mit Korrektur – 14; Sehvermögen jetzt 5/25 bds.). Die Kurzsichtigkeit wurde erst nach einem längeren Krankenhausaufenthalt im 5. Lbj. bemerkt. Nach Angaben der Mutter hat S. mit 17. Mon. gehen und mit 2 J. sprechen gelernt. Sie ent-

wickelte sich normal. Zwischen dem 4. und 5. Lbj. hatte dann S. eine Hirnhautentzündung, wovon nach Meinung der Mutter „etwas zurückblieb“ (Myopie). Sie bekam angeblich keine weiteren Krankheiten mehr, auch die sonst üblichen Kinderkrankheiten blieben aus. Nach dem Besuch des Kindergartens wurde S. zunächst in die Regelschule (Grundschule für Normalsichtige) eingeschult. Dort wurde die Pb in die vorderste Bank, dicht bei der Wandtafel, gesetzt. Von Anfang an zeigten sich allerdings Schwierigkeiten beim Lesen und Rechtschreiben. Der Mutter gegenüber äußerte sich die Volksschullehrerin dahingehend, daß sie sich in der Klasse um Einzelkinder nicht kümmern könne, dies gelte auch für S. Trotz „ausreichender“ Zensuren in allen Fächern am Ende des 1. Schuljahres repetierte S. die 1. Grundschulklasse. Im Laufe des (zweiten) Schuljahres wechselte S. dann in die Sehbehindertenschule über.

#### b) Sozial-kultureller Hintergrund:

Nach Angaben der Mutter ging S. nie gern zur Schule, dies tue sie auch jetzt nicht. Das Lesen und die täglichen Verrichtungen gingen „mit der Zeit besser“, z.B. könne S. am Sonntagmorgen alleine das Frühstück zubereiten. Nur im Rechnen sei S. schwach. Von den Kindern ihrer Nachbarschaft wird S. seit dem Besuch der Sehbehindertenschule (die in der Volksmeinung oft mit „Lernbehindertenschule“ verwechselt wird, so auch hier) immer mehr isoliert; sie gehe ja jetzt in eine „Sonderschule“. S. fährt täglich mit dem Bus zur Schule. Ihre neue Schulfreundin kann sie nur per Straßenbahn erreichen. Nach Angaben des Klassenlehrers sind die häuslichen Verhältnisse geordnet. S. fände viel Verständnis bei den Eltern. Die Mutter leidet wie die Pb an Kurzsichtigkeit, verbunden mit Nachtblindheit.

Der Klassenlehrer schätzt die Rechenleistungen der Pb besser ein als ihre Lese- und Rechtschreibleistungen (vergleiche dagegen obige Aussage der Mutter!). Im ersten Halbjahr des zweiten Schuljahrs – in der repetierten 1. Klasse – erhielt S. folgende Zeugnisnoten: Religion und Leibesübungen jeweils 2; Heimatkunde, Rechnen, Musik, Bildh. Gestalten und Text. Werken jeweils 3; Deutsch 4.

#### 3. Verhaltensbeobachtungen:

S. ist von sehr zarter Konstitution. Sie wirkt verschlossen (verängstigt?, d. Verf.) und blaß, ziemlich energielos.

Am Ende des zweiten Schuljahrs hat S. immer noch beträchtliche Schwierigkeiten mit dem Rechtschreiben und dem Lesen, deren Ursachen sich der Klassenlehrer nicht erklären kann. Auch im vorbereiteten Diktat würden der Pb überdurchschnittlich viele Fehler unterlaufen. Ähnlich sei es beim Lesen, wo die Pb „häufig“ und „meist falsch“ rate. In ihrem gesamten Verhalten werde S. immer zurückhaltender, z.B. werde ihre Stimme beim Lesen bzw. fortwährenden Leseversuch stetig leiser. Dabei scheine sie ihre „letzte Energie“ zur Überwindung der (Lese-) Schwierigkeiten zu verbrauchen. Auch mache sie einen ängstlichen Eindruck.

Der Testleiter beobachtete während der INS-Untersuchung, daß die Pb anfangs die Aufgaben sehr reserviert angegangen hat, sich aber zunehmend gelockter und interessierter gab. Sie begann mit der Lösung der Aufgabe Nr. 38 – altersentsprechend wäre ein Beginn mit der Nr. 42 – und löste als letzte richtig die Nr. 55. Zahlennachsprechen fiel ihr sehr schwer; ebenso bereiteten ihr das Aufzählen von Monatsnamen, Tieren usw. sowie Widersprüche-Finden (Spazierstock, Geldbrief) Mühe.

Während der HAWIK-Untersuchung war S. zunächst beim Verbalteil sehr zurückhaltend. Die Antworten kamen ohne langes Zögern, meist in der Form von Einwortsätzen. Hingegen „blühte“ S. im Handlungsteil förmlich auf. Dabei erzählte sie dem Testleiter, daß sie das lieber mache als Rechnen. Überhaupt fände sie Rechnen, Lesen und Schreiben langweilig. Rechnen gefalle ihr am allerwenigsten in der Schule, am ehesten noch Lesen. Bei BO und FL bedauert sie lebhaft, daß der Test schon zu Ende sei. Sie würde gerne noch weitermachen.

#### 4. Ergebnisse der Testuntersuchung mit dem INS und HAWIK (Frühjahr 1971):

Testalter: 9;6 J.

INS: Gesamt-IQ 96

HAWIK-Verbalteil: Subtests	1. AW (Allgem. Wissen)	7 WP
	2. AV (Allgem. Verständnis)	11 WP
	3. RD (Rechn. Denken)	4 WP
	4. GF (Gemeinsamkeitenfinden)	11 WP
	5. WT (Wortschatztest)	10 WP
	5a. ZN (Zahlennachsprechen)	7 WP
	Verbalpunkte:	50 / 42 WP *
HAWIK-Handlungsteil:	6. ZS (Zahlensymboltest)	7 WP
	7. BE (Bilderergänzen)	8 WP
	8. BO (Bilderordnen)	13 WP
	9. MT (Mosaiktest)	10 WP
	10. FL (Figurenlegen)	13 WP
	Handlungspunkte:	51 WP
	Gesamt-Wertpunkte:	93 WP *
Verbal-IQ:	90 *	
Handlungs-IQ:	101	
Gesamt-IQ:	95 *	

Anm.: \* = nach Umrechnung der 6 Verbaltests auf der Basis von 5 Subtests

### 5. Gutachten:

Die Frage, ob Sabines Leistungsschwäche einer generellen Lernbehinderung oder partiellen Störungen (Legasthenie) entspringt, ist nicht leicht zu beantworten, zumal sich die Symptome teilweise decken. Auch im vorliegenden Falle sind die Symptome mehrdeutig. Für eine **Lernbehinderung** spricht die schlechte Konzentrationsleistung im HAWIK (RD, ZS, BE) und INS (Beob.), die extrem niedrige Rechenleistung (RD) sowie die ausgeprägte Schwäche im HAWIK-Schulleistungsfaktor (sehr niedrige Leistungen in RD, ZN, ZS und AW.).

Gegen eine – generelle – Lernbehinderung sprechen vor allem drei, allerdings sehr gewichtige, Symptome: erstens das immerhin *durchschnittliche intellektuelle Leistungsniveau* der Probandin in den Tests (95 IQ-Punkte bzw. 96 IQ-Punkte in der HAWIK-Gesamtleistung bzw. im INS; 10 WP = 100 IQ-Punkte im MT) sowie ihre mittlere, nicht unterdurchschnittliche *Leistungskapazität* im kognitiven Bereich (z.B. 11 WP = 105 IQ-Punkte im GF); zweitens die – signifikante – *Differenz zwischen Verbal-IQ und Handlungs-IQ* zugunsten der Handlungsintelligenz (HAWIK); drittens die *überdurchschnittlich gute Pb-Leistung beim Bilderordnen* (BO im HAWIK), die Schwachsinnige kaum zustande bringen. Fast noch diskrepanter als die Testbefunde sind die Anamnese- und Beobachtungsdaten (vgl. z.B. das Kontrastsyndrom bei Anamn. a bzw. b).

Für eine vorliegende oder zumindest dominierende *legasthenische* Störung sprechen folgende Testbefunde: erstens die Differenz zwischen H-IQ und V-IQ zugunsten der HAWIK-Handlungsintelligenz; zweitens das HAWIK-Profil (Ähnlichkeit mit dem HAWIK-Profil von 92 Legasthenikern nach Klassen 1971, S. 216; vgl. auch unten S. 205f.); drittens eine ausgeprägte verbale Schwäche (AW); viertens schlechte Merkfähigkeit bzw. niedrige Gedächtnisleistungen (ZN, ZS) und fünftens eine ausgeprägte Schwäche im visuellen Erkennen und Identifizieren bekannter Figurationen (BE). Zwar liegt die Vermutung nahe, daß die Minderleistungen in BE (Bilderergänzen) zulasten der Sehbehinderung der Probandin gehen, doch widerspricht das Leistungsprofil im HAWIK-Handlungsteil (vgl. u.a. die guten Leistungen in BO und FL) einer solchen Erklärungshypothese im vorliegenden Falle. Die Schwäche in der optischen Differenzierfähigkeit bzw. Fähigkeit, Wichtiges und Unwichtiges zu unterscheiden, wirkt sich nicht nur in der Perzeption, sondern auch auf die Begriffsbildung verhängnisvoll aus. Freilich, auch die erwähnten Symptome, die eher für als gegen eine Legasthenie sprechen, sind mehrdeutiger als dem Diagnostiker lieb ist (vgl. Klasen 1971, S. 213; Schubenz & Buchwald 1964, S. 155 f. u.a.), so daß eine abschließende Urteilsfindung hier sehr erschwert wird. Trotzdem wagen wir folgende – vorläufige – Diagnose.

Folgt man der Schubenzschen Ätiologie, wonach Legasthenie Ausdruck einer zentralen Speicherschwäche ist, dann kann man im vorliegenden Falle eine *primäre Lese-Recht-schreib-Schwäche* annehmen. Diese scheint allerdings nicht unbeeinflusst zu sein von

anderen Störungen, wobei sich hier neurotische Verdachtsmomente in den Vordergrund drängen. Jedenfalls sprechen mehrere Anzeichen für eine stärkere Affektlabilität resp. (neurotische) Unsicherheit der Probandin (vgl. RD, ZN, ZS sowie Anamn. u. Beob.). Dagegen ist eine hirnorganische Schädigung mit ziemlicher Sicherheit auszuschließen (Organikersyndrom im HAWIK o.B.). Ebenfalls sprechen die gewichtigeren Untersuchungsbefunde *gegen* eine Lernbehinderung im traditionellen Sinne (siehe oben). Auch scheint die vorliegende Sehbehinderung von sekundärer Bedeutung zu sein im Hinblick auf die Spezifität des Leistungsbildes in den Testbatterien. Trotzdem sollte Sabine – schon wegen der Progressivität ihres Augenleidens – weiterhin in der Sehbehindertenschule unterrichtet werden.

Zur Behandlung der vermuteten legasthenischen Störung einerseits und der Verhaltensauffälligkeiten (neurotischen Symptome) andererseits ist hier eine spezielle Therapie unter fachpsychologischer Leitung vonnöten. Diese wird zweckmäßigerweise von einer schulpsychologischen oder Erziehungs-Beratungsstelle – in Koordination mit dem Klassenlehrer – durchgeführt. Bei dieser Gelegenheit sollten auch der Kontakt zum Elternhaus verstärkt und der positive Einfluß bedeutsamer Bezugspersonen (Lehrer, Vater?, Mutter, Geschwister?, Freundin) im Therapieplan berücksichtigt werden. Bei der Erstellung eines adäquaten Behandlungskonzeptes für Sabine wird man ohnehin auf weitere Untersuchungen, z.B. Fehleranalysen beim Lesen und Schreiben, und ggf. therapiebegleitende Diagnosen angewiesen sein.

An Hand der teilweise ausführlich beschriebenen Fallbeispiele dürften unsere theoretischen Erörterungen zur Erstellung eines Untersuchungsberichtes resp. Gutachtens nach den wichtigsten Dimensionen hin konkretisiert worden sein. Es bedarf hier wohl kaum mehr des Hinweises, daß jeder diagnostisch Tätige zu strenger Diskretion verpflichtet ist. Diese ist natürlich auch im Hinblick auf unsere Fallbesprechungen zu wahren. Wer trotz gewisser Vorkehrungen, z.B. Namensänderungen, Weglassen bestimmter Daten (Name des Testleiters, Datum der Untersuchung, genauere Kennzeichnung des Auftraggebers u.ä.), Abkürzungen u. dgl.m., einzelne Probanden wiederzuerkennen glaubt, unterliegt absoluter Schweigepflicht. Aus Gründen der Echtheit verwendeten wir keine „konstruierten“, sondern Fallbeschreibungen aus der pädagogisch-diagnostischen Praxis. Weitere diagnostische/prognostische Übungsbeispiele – unter Betonung *test*diagnostischer Aussagen – bringen wir im empirischen Teil des Buches.

Wie die bisherigen Fallbeschreibungen zeigen, kann die eigentliche Stellungnahme zum jeweiligen Untersuchungsproblem (Gutachten) länger oder kürzer, unter Umständen in *einem* Satz gefaßt (vgl. Fall E oder F), ausfallen. Die Länge oder Kürze einer Gutachtenformulierung wird in erster Linie von der Problematik des Falles und vom Auftraggeber des Gutachtens abhängen. In jedem Falle gilt die Forderung, den Untersuchungsbericht bzw. das Gutachten so *eindeutig* wie möglich abzufassen. Das Postulat der Eindeutigkeit bezieht sich nicht nur auf inhaltlich adäquate, sachlogische Formulierungen, sondern auch auf das Verständnis des Auftraggebers oder (allgemeiner) des potentiellen Lesers. Zu diesem Punkt informiert ausführlich Hartmann (1970, S. 74 ff.). Hier sei abschließend nur noch ein Problem angesprochen: die Frage der *Übermittlung* eines Gutachtens. Diese Frage hängt ja eng mit den angedeuteten semantischen Problemen zusammen.

Die Meinung darüber, ob das fertiggestellte Gutachten dem Auftraggeber (ggf. dem Probanden) *schriftlich* oder *mündlich* übermittelt werden solle, gehen in der Fachliteratur auseinander. Nach Hartmann (S. 93 f.) stehen dem Psychodiagnostiker prinzipiell vier Möglichkeiten offen, seine Befunde resp. Empfehlungen dem Auftraggeber mitzuteilen. „Er kann: 1. ein schriftliches Gutachten verfassen und auf eine mündliche Besprechung mit dem Auftraggeber verzichten; 2. ein mündliches Beratungsgespräch führen und von der Erstattung eines schriftlichen Gutachtens absehen; 3. ein schriftliches Gutachten erstellen und zusätzlich ein Beratungsgespräch anbieten oder zur Bedingung machen; 4. ein mündliches Beratungsgespräch führen und anschließend ein schriftliches Gutachten nachreichen.“



Jede Form der Gutachtenübermittlung hat ihre Vor- und Nachteile. Indessen hat sich in der diagnostischen Praxis eine *Kombination von schriftlichem Gutachten und (mündlichem) Beratungsgespräch* als der vorteilhafteste Modus erwiesen. Hartmann selbst hält die erstgenannte Möglichkeit für die schlechteste und die zuletzt genannte für die risikoärmste Form des Informationsaustausches.

„Gibt man dem Auftraggeber zuerst ein *schriftliches Gutachten* in die Hand, so kann er sich (in groben Zügen) über die Ergebnisse der Untersuchung, die Schlußfolgerungen des Diagnostikers und die eigenen Handlungsmöglichkeiten informieren. Er kommt dann bereits mit einem Vorwissen zum Beratungsgespräch. Das mag in manchen Fällen und bei manchen Auftraggebern ein Vorteil sein. In anderen Fällen – und ich nehme an, daß sie häufiger sind – dürfte diese Kommunikationssequenz gewichtige Nachteile haben. Die vielzitierten Mißverständnisse und enttäuschten Erwartungen des Gutachtenempfängers können alte Fehleinschätzungen und -einstellungen verfestigen und/oder neue ‘Vorurteile’ schaffen, die im anschließenden Beratungsgespräch mühsam wieder korrigiert werden müssen, falls das dann überhaupt noch möglich ist. Denn nun hat es der Gutachter mit Vorurteilsballast aus (mindestens) zwei Perioden zu tun. Ist er zudem genötigt, wenn nicht sein Gutachten, so doch dessen Auffassung durch den Empfänger richtigzustellen, dürfte mitunter seine Glaubwürdigkeit leiden. Denn viele Auftraggeber werden eher Widersprüche zwischen dem schriftlichen Gutachten und seiner mündlichen Auslegung wahrnehmen als eigene Mißverständnisse und Vorurteile.

Läßt dagegen ein Diagnostiker ... seinen Auftraggeber im *mündlichen Gespräch* gleichsam an der Interpretation von Untersuchungsergebnissen, an der Problemanalyse und an Entscheidungsprozessen bis hin zur Formulierung von Verhaltens- und Handlungsmöglichkeiten aktiv teilnehmen, so wirkt ein anschließend verfaßtes Gutachten wie ein gemeinsames *Kommunique*, das den Effekt des Beratungsgesprächs zu einem späteren Zeitpunkt positiv verstärken kann” (S. 97 f.).

Erst recht vorsichtig sollte man bei der gutachtlichen Aufklärung des Probanden selbst sein. Freilich wäre es gerade im Bereich der *pädagogischen* Diagnostik oft verfehlt und auch unerwünscht, dem Probanden (Schüler) das Ergebnis der (Test-)Untersuchung vorzuenthalten. Stehen Probleme der Intelligenzdiagnostik an, etwa Fragen der Schulleistungsprognose bzw. Schullaufbahnberatung, der Schulleistungsanalyse bzw. Beratung bei Lernschwierigkeiten u.ä., so wird man in der Regel keine größeren Schwierigkeiten bei der Ergebnisübermittlung haben. Jeder Schüler, auch der (in der Schule) leistungsschwächste, hat seine Stärken und positiven intellektuellen Schwerpunkte, mit deren Erörterung ein erfahrener Pädagoge und psychologisch fühlender Diagnostiker beginnen wird. Andererseits müssen natürlich die Verständnisbereitschaft und das (intellektuelle) Vermögen des Probanden, also Alter und Schulstufe, Intelligenzniveau und sozio-kulturelles Milieu des Schülers, in Betracht gezogen werden, ohne daß wir hier für jede Situation eine gültige Maxime aufstellen können. Ein sthenischer, dazu sehr intelligenter, aber „fauler“ Schüler mit erheblichen Leistungsreserven – bei aktuellen Schulleistungsschwächen – wird in der Beratung anders anzusprechen sein als ein schüchterner, „fleißiger“, aber unterdurchschnittlich begabter Schüler, der sein Leistungsmaximum bereits annähernd erreicht hat, damit jedoch seine oder seiner Eltern Bildungsambitionen kaum erfüllen kann. Die Entscheidung, auf welche Weise (wie) und wieviel (was) der Diagnostiker dem Schüler als Probanden von den Untersuchungsbefunden mitteilen kann und darf, bleibt somit letzten Endes der *pädagogischen* Verantwortung des Gutachters überlassen. Entscheidungen dieser Art werden auch dem Lehrer in der Rolle des Diagnostikers nicht erspart, womit die Funktion psychologischer Tests oder ähnlicher Untersuchungsverfahren als Unterrichts- und Erziehungshilfen im Bereich der Schule und Sonderpädagogik abschließend noch einmal bekräftigt wäre.

**FORSCHUNGSBEFUNDE ZUR BEGABUNGSDIAGNOSTIK**

**IN DER SCHULE UND SONDERPÄDAGOGIK**

# I. Untersuchungsziele

Der Intelligenzdiagnostische Anwendungsbereich ist sehr breit und vielseitig. Im Dienste klinischer, schulischer, sonderpädagogischer oder Guidance-Bemühungen überhaupt treten nachfolgend berücksichtigte Aufgabenziele in den Vordergrund:

1. Probleme der *Intelligenzdiagnose* im engeren Sinne, der *Begabungsdifferenzierung* (z.B. für Unterrichtszwecke), der *Schuleignungsermittlung* und der *Schulleistungsanalyse* (sog. Problemfälle) auf der Primar- und Sekundarstufe des *allgemeinbildenden* Schulwesens einschließlich der *Bildungsberatung* im übergreifenden Sinne.
2. *Intelligenzdiagnostische* Probleme der aufgezeigten Art im Kontext sonderpädagogischer und klinischer Fragestellungen, nämlich
  - a) Probleme der *Lernbehinderten*diagnostik;
  - b) Probleme der *Hörgeschädigten/Sprachbehinderten*diagnostik;
  - c) Probleme der *Sehgeschädigten*diagnostik..

Unsere Erörterung neuerer Untersuchungsbefunde zu den genannten Anwendungsbereichen beschränkt sich weitgehend auf die Darstellung eigener Forschungsergebnisse, wobei desiderablen resp. im Hinblick auf die praktische Notwendigkeit vordringlich erscheinenden Dateninformationen der Vorrang eingeräumt werden soll.

Die Behandlung allgemein- und sonderpädagogischer Probleme der Intelligenzdiagnostik dürfte hierbei kein Nachteil sein. Nach unseren Erfahrungen weiß gerade der an der Regelschule tätige Lehrer und Pädagoge viel zu wenig über grundlegende Probleme der sonderpädagogischen (Intelligenz-)Diagnostik Bescheid, wie umgekehrt diese ohne ständigen Kontakt mit der Entwicklung auf dem Gebiet der Psychodiagnostik allgemein sich der Gefahr des Provinzialismus ausliefert. Nach einer Aufstellung, die H. von Bracken 1965 veröffentlichte, gibt es in der BRD zur Zeit schätzungsweise 715162 behinderte Kinder und Jugendliche, d.h. rd. 11,5 % eines Altersjahrgangs müßten demnach in irgendeiner Hinsicht *sonderpädagogisch* (speziell) betreut und unterrichtet werden – und dies keineswegs nur außerhalb der Regelschulen. Die im folgenden wiedergegebenen Jahrgangsquoten (loc. cit., S. 15) basieren auf Erhebungen im Schuljahr 1962/63.

Blinde	0,013 % =	812	(abs.Zahl.)
Gehörlose	0,09 % =	5600	”
Sehbehinderte	0,1 % =	6250	”
Hörbehinderte (Schwerhörige?)	0,25 % =	15600	”
Legastheniker	0,5 % =	31200	”
Lebenspraktisch Bildungsfähige	0,5 % =	31200	”
Körperbehinderte	0,5 % =	31200	”
Sprachbehinderte	1,5 % =	93700	”
Erziehungsschwierige	2,0 % =	124900	”
Lernbehinderte	6,0 % =	374700	”

Summe der Behinderten (Anteil am Jahrgang) 11,45 % = 715162 (abs.Zahl.)

Zieht man hiervon die Anteilsquoten der lebenspraktisch Bildungsfähigen (0,5 %), der Körperbehinderten (0,5 %) und der Erziehungsschwierigen (2,0 %) ab, für die wir keine Beiträge liefern können, dann bleiben immerhin noch rd. 8,5 % (=527862) Behinderte der verschiedensten Kategorien übrig, die in den folgenden Kapiteln (Testmethoden versus Untersuchungsergebnisse) mehr oder weniger eingehend – neben den Populationen der Regelschule (Schülern an Grund-, Haupt- und Realschule sowie Gymnasiasten) – Beachtung finden werden. Aus verschiedenen äußeren (z.B. arbeitsfeldabhängigen) und inneren (d.h. sachlich-inhaltlich motivierten – siehe unsere früheren Ausführungen auf Seite 35ff.) Gründen wurde hierbei die Gruppe der Hör-/Sprachgeschädigten in gewisser Hinsicht favorisiert; gerade für diese Gruppe fehlen weithin noch angemessene und einigermaßen repräsenta-

tive Intelligenztestnormen. Um diesem Mangel wenigstens teilweise abzuhelpfen, wurden einschlägige und wohl erstmals in diesem Umfang veröffentlichte *Testnormen für Hörsprachbehinderte* in einem gesonderten *Anhang* diesem Band beigelegt. Diese sind möglicherweise auch für psycho- und soziolinguistische Forschungszwecke resp. für begabungspsychologische Forschungen überhaupt (in und außerhalb der Regelschule) von Bedeutung, womit erneut das *interdisziplinäre Kooperationsanliegen* – sowohl auf dem Felde der Forschung als auch auf praktischen Gebieten – angesprochen wäre. Dieses Postulat seiner Erfüllung ein Stück näher zu bringen, wurde bereits im Vorwort als ein Hauptanliegen dieser Veröffentlichung betont.

## II. Methodenverzeichnis (Intelligenz-/Leistungstests, Persönlichkeits-/Interessentests)

Die Auswahl der nachstehend aufgeführten Testmethoden erfolgte unter drei Gesichtspunkten. Erstens sollten die Intelligenzphänomene i.w.S. möglichst umfassend berücksichtigt werden, d.h. Tests zur Intelligenzmessung müssen sowohl die *Kernintelligenz* als auch die Sekundärintelligenz, die sog. *Hilfs- und Stützfunktionen der Intelligenz* sensu Mierke erfassen versus ein bestimmtes Faktorenmodell adäquat abdecken. Die Erfassung der Sekundärintelligenz würde beispielsweise Tests zur Messung der Leistungsmotivation, der Konzentration, des Arbeitsverhaltens oder auch (Persönlichkeits-)Interessentests einschließen. M.a.W.: Wer Begabungsphänomene diagnostisch versus prognostisch angeht, darf bei der Zusammenstellung seiner Intelligenzbatterie die Bandbreite seiner Erhebungsmethoden (Tests) nicht zu eng wählen, so daß – je nach Untersuchungszweck bzw. Zielfragen – ein möglichst repräsentatives Diagnostikum (im Hinblick auf die bezielten Merkmals- oder Fähigkeitskomplexe) zur Verfügung steht (vgl. auch Kap. I im ersten Teil dieses Bandes). Das zweite Auswahlkriterium bezieht sich mehr auf die (potentiellen) Untersuchungs*populationen*, hier also die Populationen der Regelschule auf der Primar- und Sekundarstufe sowie bestimmter Sonderschuleinrichtungen (siehe oben). Drittens wurden nur solche Testmethoden hier aufgeführt, die wir selbst erproben konnten und *mit deren Hilfe* die unten referierten *empirischen Dateninformationen gewonnen* wurden. Es leuchtet ohne weiteres ein, daß ein nach diesen Kriterien erstelltes Methodeninventar keinen Anspruch auf Vollständigkeit erheben kann (was auch nicht unserer Intention entsprochen hätte); gleichwohl dürfte für viele Untersuchungs- und Populationsbereiche eine gewisse Repräsentanz zur Zeit verfügbarer Tests bzw. tauglicher Verfahren zur Erfassung der Intelligenz oder/und relevanter Begabungsmerkmale erreicht worden sein.

Je nach der theoretischen Bezugsbasis (z.B. der g-Faktorhypothese oder einem multifaktoriellen Modellansatz) resp. der Enge oder Weite der durch den Test repräsentierten Bandbreite erfaßter Fähigkeitsmerkmale lassen sich die im nachstehend aufgeführten Katalog verzeichneten Methoden aufteilen in 1. Tests zur Erfassung der *Allgemeinbegabung*, 2. Tests zur Erfassung *differentieller* Fähigkeiten (differentielle Intelligenztests) sowie 3. Tests und Verfahren zur Erfassung *spezieller Begabungsmerkmale* (der Kernintelligenz oder sog. Hilfs- und Stützfunktionen der Intelligenz) versus bestimmter, mehr oder weniger eng umschriebener Einzelfaktoren(gruppen) der Intelligenz, z.B. sensu Thurstone. Die Hilfs- und Stützfunktionen der Intelligenz werden hier durch den Pauli-Test, den KLT, den KVT, das Verfahren nach Bourdon oder die Subtests 14 und 15 (APR) im LPS u.ä. vertreten (Erfassung der Konzentration, der Leistungsmotivation, der Anstrengungsbereitschaft u.dgl.m.); ferner wären der BIT und PIT (Interessenskalen) dem Bereich der Sekundärintelligenz zuzurechnen. Speziellere Fähigkeitsdimensionen (mehr der Kernintelligenz zugehörig) erfassen u.a. der FTU, der WST und der MTVT.

In unserer Aufstellung differenzieren wir noch zwischen *Einzel- oder Individualverfahren* (diese sind nur im Einzelversuch durchführbar) und *Gruppenverfahren* (diese können – und werden gewöhnlich – im Gruppenversuch dargeboten; prinzipiell sind die Gruppentests jedoch auch einzeln durchführbar, was in der Regel aber unökonomisch ist). Eine Untersuchungs- oder Testgruppe sollte nicht mehr als 30 bis 35 Pbn umfassen. Bei Schüleruntersuchungen empfiehlt es sich, die Tests in der (gewohnten) Klassengruppe zu administrieren.

Die folgenden Tests und Verfahren (nicht vollstandardisierten „Tests“) sind, soweit nicht anders vermerkt, bei Beltz (Weinheim), Dr. Hogrefe (Göttingen) und Huber (Bern/Stuttgart) erschienen und können über den Buchhandel oder (die Schultests von Beltz nur) direkt von den genannten Verlagen bezogen werden. Die übrigen, im Ausland publizierten, Tests werden zweckmäßig über die Testzentrale (TZ) in 7000 Stuttgart-50, Postfach 501105, Wildbader Str. 4, angefordert. Bei der Bestellung bzw. beabsichtigtem Kauf von Tests für schulische und/oder pädagogisch-diagnostische Zwecke empfiehlt sich die Legitimation durch Schul- oder Amtssiegel sowie eine kurze Angabe des Untersuchungszweckes, z.B. Schuleignungsermittlung, Schulleistungsanalyse usw. Psychologische Tests sind gesetzlich geschützt und vor Mißbrauch zu bewahren.

## 1. Tests zur Erfassung der Allgemeinbegabung

### a) Individualtests

Der **Hamburg-Wechsler-Intelligenztest für Erwachsene** (HAWIE) in der deutschen Bearbeitung von A. Hardesty & H. Lauber (Huber Verlag)

*Testaufbau:* David Wechsler, als klinischer Psychologe am Bellevue-Hospital in New York tätig, veröffentlichte 1939 die Wechsler Bellevue Intelligence Scale (WB), der 1944 die Wechsler Bellevue Adult Intelligence Scale (WAIS) folgte. Der Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE) stellt die deutsche Bearbeitung der WAIS durch A. Hardesty und H. Lauber dar und wurde 1956 von C. Bondy herausgegeben.

Der HAWIE besteht aus einem *Verbalteil* und einem *Handlungsteil*. Die 5 bzw. 6 (mit dem zusätzlichen Subtest „Wortschatz“) Unterteste des V-Teils und die 5 Unterteste des H-Teils ergeben zusammen den Gesamtest, der demnach folgende 10 bzw. 11 Subtestreihen enthält:

#### *Verbalteil*

1. Allgemeines Wissen (AW)
2. Allgemeines Verständnis (AV)
3. Zahlennachsprechen (ZN)
4. Rechnerisches Denken (RD)
5. Gemeinsamkeitenfinden (GF)
6. Wortschatztest (WT)

#### *Handlungsteil*

1. Zahlensymboltest (ZS)
2. Bilderordnen (BO)
3. Bilderergänzen (BE)
4. Mosaiktest (MT)
5. Figurenlegen (FL)

*Erfassungsdimensionen:* Da Wechsler zunächst ein differentiellles Intelligenzkonzept für diagnostische Zwecke ablehnte (vgl. seine Intelligenzdefinition oben S. 10), sollte der HAWIE (WAIS) vorab die *allgemeine Intelligenz* sensu Spearman et al. erfassen. Gleichwohl werden den einzelnen Subtests – bis auf den heutigen Tag – teilweise verschiedene Fähigkeitsbereiche zugeordnet (siehe u.a. Bondy 1956a u. 1956 (Textband) bzw. Wechsler 196 und Priester 1964). Die Korrelationen zwischen den einzelnen Untertests im HAWIE und dem Gesamtergebnis fallen unterschiedlich hoch aus. Die Interkorrelationen der Subtests sind insgesamt relativ hoch – sowohl im Intertestvergleich als auch in bezug auf den Gesamtest im Mittel um 0.68 – und sprechen somit für eine starke Sättigung auf dem g-Faktor; die Korrelation zwischen V-IQ und H-IQ beträgt sogar 0.82. Den einzelnen Subtests werden nun folgende Dimensionen zugesprochen:

- AW** Einblick in die Wissensbreite einer Person; Aufgeschlossenheit der Umwelt gegenüber; Genauigkeit und Ökonomie des Denkens; Interessensrichtungen. – Die Bearbeitung von AW-Aufgaben setzt normale Fähigkeiten voraus, verbale Kenntnisse zu erwerben; in der AW-Leistung kann sich der persönliche und familiäre „Bildungsstandard“ widerspiegeln, d.h. die Leistung in AW ist stärker vom sozio-kulturellen Milieu abhängig. – Rigide Persönlichkeiten weisen in AW überdurchschnittlich hohe Leistungen auf.  
Altersbeständigkeit: positiv (d.h. die AW-Leistung ist gegenüber altersmäßigem Intelligenzabbau verhältnismäßig resistent)  
Korrelation mit dem Gesamttest:  $r = 0.79$  (hoch).
- AV** Test des „gesunden Menschenverstandes“; situatives Verständnis, d.h. Einsicht in Feldzusammenhänge; praktische Urteilsfähigkeit; Aufgeschlossenheit; Plandenken; Fähigkeit, frühere Erfahrungen auszuwerten und zu verbalisieren. – Bei affektiven Störungen häufiger niedrige Leistungen in AV, die auch von sozialer Reife und emotionalen Einstellungen abhängig sind.  
Altersbeständigkeit: positiv  
Korrelation mit dem Gesamttest:  $r = 0.70$  (mittel)
- ZN** Kurzzeitgedächtnis oder Merkfähigkeit; Aufmerksamkeit im Sinne mühelosen, mehr passiven Registrierens äußerer Reize (im Gegensatz zur Konzentration!); Einstellungs- und Umstellfähigkeit. – Schwachsinnige erzielen oft hohe ZN-Leistungen (nur im *vorwärts* Zahlennachsprechen), während Neurotiker charakteristischerweise öfters bessere Leistungen im Zahlennachsprechen *rückwärts* als vorwärts (im intraindividuellen Vergleich) erzielen. – Auffällig niedrig fallen die ZN-Leistungen (vorwärts und rückwärts) bei hirnorganischen Störungen d. Pb aus.  
Altersbeständigkeit: negativ (d.h. in hohem Maße vom altermäßigen Intelligenzabbau betroffen)  
Korrelation mit dem Gesamttest:  $r = 0.55$  (niedrig)
- RD** Formallogisch-adäquates Denken; rechnerisches Denken (ohne anspruchsvollere rechnerische Kenntnisse); Abstraktionsvermögen auf Operationsbasis; geistige Wendigkeit; Konzentrationsvermögen (ZN und ZS mitbeachten!). – die RD-Leistung ist häufig stärker abhängig von affektiven Erfahrungen (z.B. in der Schule).  
Altersbeständigkeit: eher negativ  
Korrelation mit dem Gesamttest:  $r = 0.75$  (mittel)
- GF** Abstraktlogisches Denken; Abstraktionsfähigkeit; Denkkonsequenz und Genauigkeit; schlußfolgerndes Denken; begriffliches Denken im Verbalen bzw. verbale Begriffsbildung. – Nach Wechsler ist die GF-Leistung ein hypothetisches Maß der Intelligenzkapazität (im Gegensatz u.U. zur aktuellen Leistungsfähigkeit). – Nach Rapaport (zit. bei Priester 1964, S. 231) kann an Hand der GF-Leistung noch der Entwicklungsstand des Abstraktionsvermögens d. Pb nach drei Stufen hin differenziert betrachtet werden: 1) die *konkrete* Stufe, auf der ein Pb an mehr oberflächlichen Einzelheiten bzw. anschaulichen Gegebenheiten haften bleibt (z.B. „Apfelsine und Banane haben beide eine Schale“); 2) die *funktionale* Stufe, wo das Beziehungsdenken prävalent ist (z.B. „Apfelsine und Banane – beide kann man essen“); 3) die *abstrakte* Stufe i.e.S. (z.B. „Beide sind Früchte“), auf der begrifflich operiert wird, d.h. unabhängig von Anschauungsqualitäten und/oder Beziehungsdenken (zur eigenen Person) wesentliche Gemeinsamkeiten zweier Gegenstände oder Sachverhalte hervorgehoben werden.  
Altersbeständigkeit: negativ  
Korrelation mit dem Gesamttest:  $r = 0.81$  (hoch)
- WT** Sprachliche Ausdrucksfähigkeit; allgemeines Wissen und Vorstellungsbesitz (Begriffsschatz); Denktypologische Artung geistiger Operationen; Anschaulichkeit des Denkens. – Maß der Lernfähigkeit; Bestand an sprachlichen Kenntnissen (abhängig von der Lebenserfahrung d. Pb sowie sozio-kulturellen Determinanten).

Nach Rapaport ist der WT einer der zuverlässigsten Unterteste im HAWIE; er repräsentiere die „ursprüngliche“ Intelligenz.

Altersbeständigkeit: positiv

Korrelation mit dem Gesamttest:  $r = 0.89$  (sehr hoch)

- ZS** Gedächtnisleistung; Speedfaktor im Psychomotorischen bzw. allgemeine psychomotorische Geschwindigkeit; Einbildungsfähigkeit für bildlich-sinnfreie Figurationen; Assoziationsfähigkeit (mit Symbolen); visuell-motorische Koordination; Lernfähigkeit und Konzentrationsvermögen (im Handlungsbereich). – Neurotisch Unsichere weisen öfters niedrige Leistungen in ZS auf.  
Altersbeständigkeit: negativ  
Korrelation mit dem Gesamttest:  $r = 0.63$  (niedrig)
- BO** Visualität; Fähigkeit, soziale Gesamtsituationen zu erfassen bzw. zu verstehen und auszuwerten; Produktivität; Beobachtungsgabe; Kombinatorik; Planning ability and Anticipation („Einstellung“ sensu Rapaport). – BO erfordert ähnliche Funktionen wie AV, wobei jedoch folgender Unterschied beachtet werden muß: bei BO erfolgt die Lösung mehr emotional und intuitiv, bei AV mehr rational; ferner spielt bei BO der Auffassungstyp (analytisch versus ganzheitlich) eine Rolle. – Schwachsinnige perseverieren bei der Lösung von BO gern; sofern jedoch BO gut gelöst wird, ist Schwachsinn auszuschließen. Dies gilt auch im Fall sog. Pseudodebilität, wenn beispielsweise verwahrloste Jugendliche bei insgesamt schlechten HAWIE-Leistungen in BO gut abschneiden (Kontra-Indikation einer Lernbehinderung!).  
Altersbeständigkeit: negativ  
Korrelation mit dem Gesamttest:  $r = 0.62$  (niedrig)
- BE** Visuelles Erkennen bzw. Identifizieren bekannter Figuren und Gegenstände; Perception und Begriffsbildung; Beobachtungsgabe; optische Differenzierungsfähigkeit; Fähigkeit, Wichtiges von Unwichtigem zu unterscheiden. – Test der willkürlichen, fixierenden Aufmerksamkeit (Konzentration); als Konzentrationstest u.a. geeignet zur Feststellung von Lernbehinderungen (vgl. Kap. III, 2a unten).  
Altersbeständigkeit: positiv  
Korrelation mit dem Gesamttest:  $r = 0.70$  (mittel)  
*Kritik:* BE differenziert unzureichend zwischen hoher und sehr hoher Intelligenz resp. Konzentrationsfähigkeit („niedrige Testdecke“ oder sog. „Ceiling“-Effekt)
- MT** Umstrukturierungsfähigkeit; räumliches Vorstellungsvermögen; Abstraktionsfähigkeit; Differenzierungsfähigkeit; Konstruktivität; Kombinationsfähigkeit. – Synthetische und analytische Fähigkeiten; visuell-motorische Koordination. – Allgemein wird die Leistung im MT als ein guter Maßstab für „general intelligence“ (Allgemeinbegabung) angesehen. – Sofern die MT-Leistung hoch ist und die Leistungen in den übrigen Subtests niedrig ausfallen, besteht Verdacht auf neurotische Störung(en) – Die MT-Leistung ist im allgemeinen niedrig bei Schwachsinn, bei vorzeitigem Altersabbau bzw. vorliegenden Hirnerkrankungen.  
Altersbeständigkeit: negativ  
Korrelation mit dem Gesamttest:  $r = 0.76$  (mittel/hoch)
- FL** Ekphorierfähigkeit; Klarheit und Schärfe bildlicher Vorstellungen; Art der Zielgerichtetheit; Antizipation; planende Phantasie; visuell-motorische Koordination. – Aufmerksamkeit; Sorgfaltsstreben; Arbeitshaltung.  
*Anmerkung:* Die Charakteristik des Vorgehens bei der FL-Bearbeitung sollte durch eine (zusätzliche) Verhaltensbeobachtung festgehalten werden!  
Altersbeständigkeit: positiv  
Korrelation mit dem Gesamttest:  $r = 0.54$  (niedrig). Interessanterweise steigt die Korrelation bei fortschreitendem Alter deutlich an und nähert sich Koeffizienten, die Priester im Kindesalter beim HAWIK (siehe unten) fand.

**Testgütekriterien:** Eine systematische *Itemanalyse* wurde beim HAWIE nicht durchgeführt. Lienert & Fickert (zit. nach Priester 1964, S. 234) haben nach Erscheinen des HAWIE 1958 an Hand einer – freilich nicht populationsrepräsentativen – Gymnasialstichprobe (N = 217) Schwierigkeit und Trennschärfe des verbalen Aufgabenteils kontrolliert. Dabei erwies sich der Subtest RD bei einem mittleren Schwierigkeitsindex von  $M_p = 78$  (bei Gymnasialisten) als relativ leicht, während der Subtest WT bei einem  $M_p = 47$  etwas zu schwer erscheint; unter altersspezifischem bzw. nicht schultypischem Normenbezug darf wohl eine noch größere WT-Schwierigkeit (als angegeben) angenommen werden. Lienert & Fickert fanden ferner, daß die Hälfte der Aufgaben des HAWIE-Verbalteils hinlänglich trennscharf ist; nur 17 % der Aufgaben zeigten unzureichende Trennschärfeindices.

Die *Objektivität* der Testdurchführung kann als gesichert gelten. Bei der HAWIE-Auswertung sind hingegen subjektive Einflüsse des Testleiters, z.B. in AV, GF und WT, nicht immer auszuschließen. Insgesamt befriedigt jedoch die Objektivität des HAWIE, sofern sich der Testleiter genau an die vorgeschriebene Instruktion und Auswertungskriterien (mit Beispielhilfen) hält. Dessen ungeachtet wären Untersuchungen zur Auswertungsreliabilität des HAWIE desiderabel.

Die (Split-half-) *Reliabilität* des HAWIE weist eine interne Konsistenz von 0.94 (nach der Spearman-Brown-Formel aufgewertet) bei einer Stichprobe von 370 Pbn im Alter von 20 bis 34 Jahren auf. Verbal-IQ und Handlungs-IQ korrelierten bei derselben Stichprobe mit  $r = 0.82$  hoch (siehe Erfassungsdimensionen oben); die mittlere Differenz zwischen V-IQ und H-IQ betrug bei der Standardisierungsstichprobe etwa 0 IQ-Punkte. „Die Analyse der *Differenzwerte* (zahlenmäßige Differenz zwischen dem *Verbal-* und *Handlungs-IQ* eines Prüflings) ist nach unserer Meinung von besonderer Bedeutung, da die Differenzen in der Testleistung bei beiden Testteilen wegen ihrer hohen Reliabilität zur klinischen Diagnose verwertet werden können. Dazu muß man aber vorher wissen, wie groß der Differenzwert sein muß, um eine klinische Aussage zuzulassen. Nach den bereits vorliegenden Daten des HAWIE kann gesagt werden, daß etwa 50 % einer unausgewählten Gruppe von Prüflingen IQ-Differenzen zwischen *Verbal-* und *Handlungsteil* von 8 und mehr IQ-Punkten erhalten. Diese Ergebnisse decken sich mit denen der amerikanischen Originalform und mit denen des HAWIK.“ (Priester 1964, S. 236 f.)

Die Retest-Reliabilität liegt in bezug auf den Verbalteil zwischen 0.92 und 0.95, in bezug auf den Handlungsteil zwischen 0.80 und 0.86, in bezug auf die Gesamt-Punkte zwischen 0.90 und 0.94. Ähnlich hohe, d.h. befriedigende Stabilitätskoeffizienten wurden nur noch bei den Subtests AW (0.83 - 0.91), RD (0.73 - 0.84) und GF (0.85 - 0.90) gefunden. während ZN und BO die niedrigsten (Retest-)Reliabilitätskoeffizienten überhaupt aufwiesen. „Es ist deutlich, daß die Reliabilität der Testteile (*Verbal-, Handlungs- und Gesamtteil*) befriedigend hoch ist, bei den Untertests aber mit wenigen Ausnahmen sehr zu wünschen übrig läßt. Die niedrige Reliabilität der Untertests ist vor allem bedingt durch die relativ geringe Anzahl ihrer Testaufgaben – eine unvermeidliche Konsequenz der Zeitbegrenzung bei der Intelligenzmessung. Die niedrige Reliabilität der Untertests bedingt große Vorsicht bei Interpretationen der Differenzen zwischen Wertpunkten (Profilanalyse). Es muß dabei immer beachtet werden, daß die Differenzen eher auf Meßfehler als auf wahre Differenzen der gemessenen Fähigkeiten zurückgehen“ (loc. cit.).

Der *Standardmeßfehler*, der ein Maß für die Fehlerbreite einer Testbewertung darstellt und somit bei der Interpretation von inter- oder/und intraindividuellen Testleistungsdifferenzen eine wichtige Rolle spielt (siehe oben S. 66ff.), beträgt für den *Verbalteil* ca. 3 IQ-Punkte, für den *Handlungsteil* ca. 5 IQ-Punkte und für den *Gesamt-HAWIE* ca. 7 IQ-Punkte.

Die *Validität* des HAWIE ist bislang noch nicht systematisch untersucht worden. Die vorliegenden Angaben (unterschiedlichster Provenienz) weisen Korrelationen zu einzelnen Binetskalen zwischen 0.7 und 0.8, zum Schulerfolg – je nach Schultyp bzw. Er-



folgskriterium und Pbn-Gruppe – zwischen 0.4 und 0.6 auf. Mit dem Intelligenz-Struktur-Test (IST) von R. Amthauer korreliert der HAWIE mit  $r = 0.74$ .

Faktorenanalytische Befunde zum HAWIE haben bisher nur Lienert & Fickert (1958) sowie Riegel (1960) vorgelegt. Demnach können vier Faktoren(gruppen) in der HAWIE-Leistung unterschieden werden: 1) ein *genereller (g-)*Faktor, der etwa 50 % der Gesamtvarianz einnimmt; 2) ein Faktor *Sprachbeherrschung* (verbales Verständnis). „Sprachliches Verständnis ist die Fähigkeit, den Bedeutungsinhalt von einzelnen oder kombinierten Worten zu verstehen. Dieser Faktor wird am besten im WT, AW, AV und GF repräsentiert. Nach aphasischen Störungen zu beobachtende Defekte scheinen darauf hinzuweisen, daß dieser Faktor eine kortikale Lokalisierung aufweist.“ (Priester 1964, S. 239); 3) ein *Handlungsfaktor* (nichtverbale Organisation). „Er wird besonders deutlich beim FL und beim MT. Das BO und das BE zeigen ebenfalls geringe Ladungen. Dieser Faktor scheint komplexer zu sein als das sprachliche Verständnis, was sich in den vielen verschiedenen Interpretationen verschiedener Autoren widerspiegelt. Sein Hauptbestandteil scheint die Fähigkeit zur Organisation einzelner, räumlich wahrgenommener, Einheiten zu größeren Ganzen oder Konfigurationen zu sein“ (loc. cit.); 4) ein *undifferenzierter Gedächtnisfaktor*, von manchen Autoren auch als *Rechenfertigkeit* interpretiert. „Die höchsten Ladungen mit ihm zeigen ZN, ZS und bei einigen Altersgruppen RD und AW. Dieser Faktor repräsentiert nach Wechsler eine Art allgemeines Behaltenkönnen (general retentiveness), das operational wohl am besten als assoziatives Gedächtnis identifiziert werden kann. Interessanterweise steigt die Faktorenladung mit diesem Faktor mit zunehmendem Alter deutlich an. Der ältere Mensch ist mehr und mehr abhängig vom früheren Erfahrungsgut (aufgespeichertes Wissen) und weniger von primären Fähigkeiten. Sobald ein gutes Gedächtnis ihn verläßt, ist nicht mehr viel vorhanden“ (loc. cit.). Schließlich fanden Lienert & Riegel noch eine Reihe zusätzlicher (spezifischer) Faktoren, die jedoch weniger als 2 % der Gesamtvarianz ausmachen, somit die *Faktorenstruktur* des HAWIE nur unwesentlich beeinflussen.

In der Testliteratur finden sich immer wieder Hinweise für eine syndromatische Ordnung der HAWIE-Befunde auf der Basis der – zweifelhaften – Profilanalyse. Am häufigsten werden folgende genannt.

**Organikersyndrom:** niedrige HAWIE-Leistungen in ZS, FL und MT.

**Schwachsinnigensyndrom** (Lernbehinderung): sehr niedrige bzw. (intraindividuell) niedrige Leistungen in RD, ZS und AW (Faktor bzw. Faktorengruppe der „Lernfähigkeit“); vgl. HAWIK unten sowie Kap. III, 2a.

**Aphasiesyndrom:** niedrige HAWIE-Leistungen in WT, AW, AV und GF (siehe oben).

**Oberschulsyndrom** (Gymnasialeignung): hohe Leistungen in AW, RD, GF und WT.

Bei der Verwendung dieser Indikationen sollte man stets die (Un-)Reliabilität der betr. Subtests berücksichtigen (vergleiche auch den nachstehenden Punkt „Kritik“ des HAWIE).

Die *Normen* des Wechsler-Tests sind für die einzelnen Subtests in WP-Einheiten (Wertpunkten) dargestellt, die durch einfache lineare Transformation aus den Rohpunkten gewonnen wurden ( $M_{WP} = 10$ ;  $s_{WP} = 3$ ). Sie basieren auf den HAWIE-Leistungen von 370 Pbn im Alter von 20 bis 34 Jahren der Standardisierungsstichprobe. Die WP-Summen des Verbal-, Handlungs- und Gesamttests ergeben – durch IQ-Transformation – den *Verbal-IQ*, *Handlungs-IQ* und *Gesamt-IQ*. „Zur Aufstellung der drei IQ wurden bei allen Altersstufen (Alter 10 – 59) die Wertpunktverteilungen ermittelt und deren statistische Konstanten berechnet. Die Ergebnisse zeigten, daß die Mittelwerte der Wertpunktsummen mit dem Alter fortschreitend bis zu etwa 28 Jahren ansteigen, um dann deutlich wieder abzufallen. Dieser Anstieg der Mittelwerte bis zum Alter 28 steht im Widerspruch zu den Normen anderer Tests, ist aber bisher unseres Wissens noch nicht interpretiert worden. Bei der neuen Standardisierung des amerikanischen Originaltests (Wechsler 1955) wurden ähnliche Ergebnisse gefunden, wobei der Anstieg der mittleren Werte von manchen Autoren (Anastasi 1961) als ein Ergebnis der breiteren mo-

dernen Erziehung angesehen wird. Nach unserer Meinung ist dieser Anstieg ebenfalls als Funktion der neuartigen Aufgabenzusammenstellung und der adäquateren Stichprobenerhebung anzusehen. Die Variationskoeffizienten der Wertpunkte sinken bei den Jugendlichen und steigen bei Erwachsenen mit dem Alter. Diese Resultate decken sich mit anderen Untersuchungen, sind aber bisher nicht eindeutig interpretiert worden. Die IQ beim HAWIE werden als *Abweichungs-Intelligenzquotienten* bezeichnet mit dem Mittelwert 100 und der Standardabweichung 15. Sie werden nicht wie bei den Altersskalen durch das Verhältnis von Intelligenzalter zu Lebensalter (IA/LA) definiert, sondern durch die individuelle Abweichung vom Mittelwert der jeweiligen Altersgruppe (erzielte oder tatsächliche Wertpunktschuld/erwartete mittlere Wertpunktschuld der Altersgruppe). Als grundlegender Vorzug der HAWIE-IQ muß die Konstanz der IQ-Variabilität bezeichnet werden. Ein bestimmter IQ besitzt immer die gleiche Bedeutung in bezug auf die relative Position eines Prüflings zu seiner Altersgruppe. Bei der Benutzung der HAWIE-IQ sollte immer beachtet werden, daß die Bewertung eines Probanden beständig mit der mittleren Leistung seiner eigenen Altersgruppe verglichen wird und nicht – wie etwa bei den *Binet-Skalen* – mit einem hypothetischen durchschnittlichen Erwachsenen-Intelligenzalter. Bei Vergleichen zwischen HAWIE- und anderen Intelligenzeinstufungen ist darauf zu achten, daß bei gleicher, konstanter oder objektiver Leistung eines Probanden der HAWIE diesen mit steigendem Alter höher einstufen wird, da seine Testbewertung mit ständig sinkenden Altersnormen verglichen wird. Wenn z.B. ein 20jähriger Proband den IQ 60 erhält, wird er bei gleicher Leistung im Alter von 50 den IQ 70 erhalten" (Priester 1964, S. 235 f.). Sämtliche Normdaten, d.h. WP-Tabelle und IQ-Tabellen (Altersnormen von 10;0 bis 59 J.) finden sich im Textband zum HAWIE (vgl. Bondy 1956 bzw. Wechsler 1961).

**Anwendungsbereiche:** Der HAWIE gehört zu den meistverwendeten (Individual-)Tests zur Messung der Intelligenz im späten Kindes-, Jugend- und Erwachsenenalter, also vom 10. Lebensjahr aufwärts bis zum 60. Lebensjahr und darüber. Seine größte Bedeutung erlangt das Diagnostikum in der psychologischen (Erziehungs- und schulpyschologischen) Beratung, in der klinischen Psychologie und in der Sonderpädagogik (vgl. Kap. III, 2a-c). Einige Spezialfragen (Profilauswertung, Abbauindex) werden uns im Zusammenhang einer *kritischen* Würdigung der Wechsler-Tests nach der Besprechung des HAWIK abschließend noch beschäftigen.

Der **Hamburg-Wechsler-Intelligenztest für Kinder** (HAWIK) in der deutschen Bearbeitung von F.P. Hardesty & H.J. Priester (Huber Verlag)

**Testaufbau:** Der HAWIK stellt die deutschsprachige Adaptation der Wechsler Intelligence Scale for Children (WISC) dar, die D. Wechsler 1949 quasi als Erweiterung und Ergänzung der WAIS nach unten, d.h. für die jüngeren Jahrgänge, herausbrachte. Die WISC bzw. der HAWIK sind deshalb im Aufbau nahezu mit der WAIS bzw. dem HAWIE identisch. Genau so wie der HAWIE besteht der HAWIK aus einem Verbalteil und einem Handlungsteil, die beide zusammen wiederum den Gesamttest ergeben. Lediglich zwei Untertests wurden im Verbalteil vertauscht. Statt des im HAWIE obligatorischen ZN und fakultativen WT wird im HAWIK der Subtest WT obligatorisch und ZN fakultativ (als Zusatztest) angeboten. Ansonsten gelten praktisch die oben angeführten Erfassungsdimensionen der 10 bzw. 11 Subtests auch hier wieder.

**Testgütekriterien:** In etwa können die beim HAWIE ermittelten Daten zur Standardisierung, Objektivität, Reliabilität und Validität des Verfahrens – sinngemäß – auf den HAWIK übertragen werden. So liegen nach Wewetz (1964, S. 210) die Interkorrelationen beim HAWIK für die Gruppe der 7-, 10- und 13-jährigen Pbn bei 0.33, 0.42 und 0.39 (beim WISC zum Vergleich: 0.34, 0.43 und 0.39). Obgleich die Faktorenstruktur des HAWIE cum grano salis auch für den HAWIK relevant ist, scheint hier doch stärker der g-Faktor zu dominieren, was ganz im Sinne der Wechsler'schen Intelligenzdefinition liegen dürfte. WISC und HAWIK wurden mit zahlreichen anderen Intelligenztests als Außenkriterien korreliert. Im allgemeinen zeigte sich eine gute bis sehr gute Übereinstimmung zu den ver-

schiedenen Binetskalen, zum California Test of Mental Maturity von W.W. Clark et al. u.ä. Tests, wobei die Validitätskoeffizienten nicht selten Werte um 0.8 und höher erreichen. Allerdings streuen die mitgeteilten Übereinstimmungskoeffizienten teilweise beträchtlich, doch lassen sich mindestens zwei Trends beobachten: (1) die *Abnahme der durchschnittlichen Übereinstimmungsvalidität bei ansteigendem Bildungsniveau* (z.B. 0.8 bei Hilfschülern versus 0.6 bei Gymnasiasten) nach einer Untersuchung von Priester (1959) mit dem HAWIK bzw. Stanford-Intelligenztest sensu Lückert. In allen drei untersuchten Schülerstichproben (Lernbehinderte, Volksschüler, Gymnasiasten) lagen bei der Mehrzahl der Fälle, besonders deutlich bei den Oberschülern, die HAWIK-IQ *über* den Stanford-IQ, und zwar im Mittel bei 66 % der Pbn. Im Gegensatz dazu zeigte sich (2) *fast bei sämtlichen (übrigen) Binet-Test- und HAWIK- bzw. WISC-Vergleichen, daß der Wechsler-Test „schwerer“ ist*, wobei sich die größten Differenzen bei hohem Intelligenzniveau manifestierten (z.B. 130 Wechsler-IQ ~ 146 Stanford-Binet-IQ). Die beobachteten Diskrepanzen mögen aber auch – wenigstens teilweise – durch meßtheoretische Voraussetzungs-differenzen zu erklären sein, worauf schon u.a. Wewetzer hingewiesen hat. Je nach Schultyp ist der HAWIE bzw. HAWIK mit dem Schulerfolg zwischen 0.4 und 0.6 korreliert.

Retest-*Realibilität* sowie interne Konsistenz des HAWIK in bezug auf die Gesamtskala sind durch Koeffizienten um 0.9 (0.91, 0.92 und 0.93 für die 7-, 10- und 13jährigen) ausgewiesen. Für die WISC wurde noch nach vier Jahren eine Retestrealibilität (Teststabilität) von  $r = 0.77$  ermittelt. Weitere Standardisierungsinformationen finden sich bei Priester (1958) sowie bei Wewetzer (1964). Die HAWIK-Altersnormen (Tabellen für 6. bis 16. Lbj.) sind der Handanweisung zum HAWIK beigegeben.

*Anwendungsbereiche:* Im wesentlichen sind hier die gleichen Anwendungsbereiche wie die oben beim HAWIE geschilderten – jetzt auf die 6- bis 15jährigen bezogen – einschlägig. Darüber hinaus empfiehlt sich der HAWIK bei der Intelligenzdiagnose schwachsinniger Jugendlicher und Erwachsener (Eyferth 1963).

*Kritische Würdigung* der Wechsler-Testreihen (HAWIE und HAWIK): Die Wechsler-Skalen gehören zu den meistverwendeten (Individual-)Intelligenztests überhaupt. Die deutschen Versionen, HAWIE und HAWIK, verfügen über eine *hinreichende Objektivität* sowie eine *gute Reliabilität* – sowohl des Gesamttests als auch seiner Teile. Weitere *Vorteile* liegen in der *sorgfältigen Standardisierung* und *guten Testeichung* einerseits und der *vielseitigen Anwendungsmöglichkeiten* andererseits. So wird der HAWIK bzw. HAWIE beispielsweise auch bei Blinden und hochgradig Sehbehinderten (Verbalteil), bei Hör- und Sprachbehinderten (Handlungsteil) oder in der Lernbehindertendiagnostik (Selektion) verwendet. Siehe auch Tab. III im Anhang dieses Buches.

Andererseits müssen mindestens drei nicht unerhebliche *Nachteile* genannt werden: 1. die mehr oder weniger *willkürliche Auswahl* der (genauer: *einzelner*) Intelligenzfaktoren und die noch sehr *lückenhafte Testvalidierung*; 2. die teilweise damit verbundene *hohe Redundanz* der Erfassungsdimensionen (siehe oben) und 3. *zu hohe Testinterkorrelationen*, die eine – von Wechsler selbst befürwortete – *Profilauswertung* fragwürdig machen. Gelegentlich hört man auch den Einwand, daß die HAWIK-Leistungen zu sehr milieuhabhängig ausfielen, was sich etwa in der *verbalen Leistungsüberlegenheit* (V-Teil gegenüber H-Teil) der Mittelschichtkinder versus der Testleistungsüberlegenheit dieser Gruppe überhaupt gegenüber Unterschichtkindern manifestiere. Diese oder ähnliche Einwände sind kaum von der Hand zu weisen, wenngleich empirische Kontrolluntersuchungen darüber nur vereinzelt vorliegen und bislang kein einheitliches Bild ermöglichen. So konnte Estes (1953 u. 1955) in einer Explorationsstudie bei zweimal 40 Schulkindern des 7. und 10. Lebensjahres entsprechende Tendenzen im WISC zunächst nachweisen; bei den drei Jahre später durchgeführten Retestungen waren jedoch die Mittelwertsdifferenzen der Mittel- und Unterschichtkinder soweit geschrumpft, daß sie nicht mehr signifikant zu machen waren. Wir wollen deshalb hier nur auf die drei zuerst genannten Punkte der Testkritik eingehen.

Ad 1.: Die Wechsler-Skalen nehmen hinsichtlich ihrer faktoriellen Erfassungsstruktur gewissermaßen eine Zwischenstellung von Binetarium als Form der traditionellen Intelligenztestmessung und modern konzipierter, meistens faktorisierter Intelligenztests ein. Diese Zwitterstellung kommt bereits in dem doppelten Anspruch Wechslers zum Ausdruck, einerseits die Allgemeinbefähigung (vergleiche die Intelligenzdefinition auf der Basis der g-Faktorhypothese) und andererseits wie auch immer umschriebene Strukturmomente der Intelligenz (Profilanalyse) zu erfassen. Unter den vier bisher faktorenanalytisch nachgewiesenen Faktoren beansprucht jedoch der g-Faktor mit Abstand den Löwenanteil (quer durch die einzelnen Subtests), weshalb wir HAWIE und HAWIK – entsprechend dem Wechslerschen Intelligenzkriterium (s.S. 10) – den Verfahren zur Messung der Allgemeinbegabung zuordneten.

Ad 2.: Die teilweise hohe Redundanz der Untertests, die sich u.a. schon in den Korrelationskoeffizienten der Subtests in bezug auf den Gesamt-IQ andeutet und eo ipso auf eine hohe interne Konsistenz des (Gesamt-)Testes hinweist (vergleiche auch die Beschreibung der Erfassungsdimensionen auf S. 114ff. und deren inhaltliche Überschneidung in den Subtests), legte die Erstellung von Kurzformen der Wechsler-Tests nahe. Nach Pawlik (1970, S. 139, vgl. auch Dahl 1968) erwies sich dabei folgende Subtestkombination als die *beste altersunabhängige Kurzfassung*: AW, AV, BO und FL. Die Ergebnisse dieser vier Unterteststrecken korrelieren mit dem Gesamt-IQ (der Normal- oder Langform) mit 0.93, mit dem Verbal-IQ mit 0.91 und mit dem Handlungs-IQ mit 0.88. Bei Verwendung dieser oder jener Kurzform des HAWIK empfiehlt es sich jedoch, die für die betr. Altersstufe optimale Subtestzusammenstellung jeweils selbst zu ermitteln bzw. ad hoc zu kontrollieren, indem man die entsprechenden Subtestresultate kombiniert mit dem G-IQ, V-IQ und H-IQ (der Normalform) korreliert. Die Kontrollstichprobe sollte im Hinblick auf die Anwendungspopulation hierbei in etwa repräsentativ zusammengestellt werden. Aufgrund eigener empirischer Untersuchungen schlug Pawlik (loc. cit.) für die Altersgruppe der *Siebenjährigen* beispielsweise *folgende Subtestkombinationen* vor:

- a) AW, RD, ZS und FL (maximale Korrelation – optimaler Zusammenhang – mit dem Gesamt-IQ = 0.95);
- b) AW und WT (maximale Korrelation mit dem Verbal-IQ = 0.85);
- c) BO und FL (maximale Korrelation mit dem Handlungs-IQ = 0.89).

Ad 3.: Zweifellos die schwerwiegendsten Einwände wurden gegen die Wechsler-Tests als Profilverfahren vorgebracht. Eine Profilauswertung für differentialdiagnostische Zwecke erscheint von der Anordnung der Subtests bzw. Testteile her intuitiv sehr verlockend, wenngleich diese unter dem Gesichtspunkt der WP-Normierung strenggenommen nur für den HAWIE und die Altersgruppe der 20- bis 34jährigen in Anspruch genommen werden dürfte. Gleichwohl sind profilanalytische Auswertungsvorschläge auch zum HAWIK bis auf den heutigen Tag nicht verstummt. Die wichtigsten seien deshalb – zusammen mit den vorgebrachten Kritiken – in aller Kürze hier erörtert.

a) Ein *Vergleich zwischen Verbal-IQ und Handlungs-IQ* ist die einfachste und wohl unproblematischste Form der HAWIE- bzw. HAWIK-Profilauswertung. So wird etwa ein Testleistungsdefizit im Wechsler-Verbalteil (V-IQ deutlich kleiner als H-IQ) fallweise interpretiert als schlechte oder inadäquate Schulbildung, als Underachievement, als extrem praktische Befähigung, als soziale Fehlanpassung, als kortikale Schädigung (der dominanten Hirnhemisphäre) u. dgl.m.; in der Regel wird sich der Stellenwert des Befundes (V-IQ kleiner als H-IQ) also erst aus dem übergeordneten Kontext ermitteln lassen (siehe auch S. 97f. in diesem Buch). Entsprechend werden Testleistungsdefizite im Handlungsteil (H-IQ kleiner als V-IQ) fallweise als stärker reduziertes Allgemeininteresse (z.B. bei Depressiven), als einseitige (Schul-)Bildungseffekte versus hypertrophierte Anpassungstendenzen u.dgl.m. gedeutet. Im Hinblick auf die *Schuleignungsermittlung* sind vielleicht noch folgende Hinweise interessant: *Gymnasiasten* weisen öfters – aber keineswegs immer! – einen höheren Verbal-IQ im intraindividuellen Vergleich zum Handlungs-IQ auf, wohin-

gegen *Realschülern* häufiger ein höherer Handlungs-IQ (als Verbal-IQ) nachgesagt wird. Nach einer Untersuchung von Pawlik zeigte eine Gruppe von Oberschulversagern ebenfalls bessere HAWIE-Leistungen im Handlungsteil im Vergleich zum Verbalteil. Solche oder ähnliche – mehr oder minder statistisch kontrollierten – Erfahrungshinweise bedeuten allenfalls gelegentlich eine wirkliche Hilfe im psychodiagnostischen Entscheidungsprozeß. Speziell bei eignungsdiagnostischen Fragestellungen wird man nach verlässlicheren Maßstäbkriterien Ausschau halten müssen (siehe dazu ausführlicher Kap. III, 1 unten sowie Heller 1970a, S. 104 ff.): insonderheit gilt dies für die Treffsicherheit von Diagnosen versus Prognosen im Einzelfall.

b) Vorab in der klinischen und/oder sonderpädagogischen Diagnostik spielt der von Wechsler vorgeschlagene *Abbauindex* (Deterioration Index) eine Rolle. Dieser Index (DI) basiert auf der Annahme, daß die einzelnen psychischen Leistungsfunktionen vom (normalen) Altersabbau unterschiedlich stark betroffen seien. „Im allgemeinen wird ein Intelligenzabbau am besten durch eine Messung der Beantwortungsgeschwindigkeit, der Lernfähigkeit und der Fähigkeit, neue, besonders räumliche Konfigurationen zu erfassen, aufgedeckt“ (Wechsler 1961, S. 81). Nach Wechsler werden vom altersmäßigen Intelligenzabbau besonders stark betroffen die Subtests – des HAWIE – ZN, RD, ZS, MT und GF (BO ist hier fraglich), während die Subtests AW, AV, FL, BE und WT relativ stabil, d.h. gegen Abbau resistent, seien. Der Verfall kann

$$\text{als Prozentdifferenz} \quad DI = \frac{\text{Sum WP stab. Tests} - \text{Sum WP instab. Tests}}{\text{Summe WP stab. Tests}} \cdot 100$$

$$\text{oder als Abbauquotient} \quad DI = \frac{\text{Sum WP instab. Tests}}{\text{Sum WP stab. Tests}} \cdot 100$$

ausgedrückt werden. „Um ein Maß des Abbaus zu gewinnen, vergleicht man die Summe der Wertpunkte der „beständigen“ Tests mit denen der „nicht-beständigen“ Tests. Dabei muß man Unterschiede in der Anzahl der Tests jeder Gruppe berücksichtigen. Für einen genaueren Vergleich stellt man die Summe der ersten vier „beständigen“ Tests den ersten vier Tests der „nicht-beständigen“ Tests. Dabei muß man Unterschiede in der Anzahl der Tests jeder Gruppe berücksichtigen. Für einen genaueren Vergleich stellt man die Summe der ersten vier „beständigen“ Tests den ersten vier Tests der „nicht-beständigen“ Gruppe gegenüber. Das Ergebnis kann entweder als Quotient oder als Differenz beider Summen ausgedrückt werden. Sofern das Ergebnis als Differenz ausgedrückt wird, muß es in eine Prozentdifferenz umgerechnet werden, um die absolute Größe der verglichenen Summen zu berücksichtigen. Wenn also die Summe der „beständigen“ Untertests einer Person 50 ist und die Summe der „nicht-beständigen“ Untertests 40, beträgt der Verlust durch Abbau 20 Prozent und der Abbauquotient 80. Hoher Verlustprozentsatz oder niedriger Abbauquotient (erheblich unter 100) zeigen Abbauerscheinungen an“ (loc. cit., S. 77 f.). Diagnostisch ist natürlich der *vorzeitige* Abbau, etwa als Symptom hirnorganischer Krankheitsprozesse, interessant. Die Verwendung des Abbauindex ist in den letzten Jahren verschiedentlich kritisiert worden, wobei der Einwand, daß Wechslers Befunde zum Abbauindex auf Querschnittstudien (statt auf hier relevanteren Längsschnitten) basierten, wohl am schwersten wiegt. Auch konnte in verschiedenen Kontrolluntersuchungen die Reliabilität und Validität der indexmäßigen Subtestanalyse nicht eindeutig gestützt werden, so daß für entsprechende Interpretationsversuche mit Hilfe des Abbauindex größte Vorsicht am Platz ist.

c) Häufig wird auch die WP-Variabilität der Subtests selbst zum Anlaß genommen, eine auf die oben beschriebenen Subtestdimensionen des HAWIE bzw. HAWIK bezogene Profilanalyse durchzuführen, obwohl ja die Wechslerkalen ursprünglich nicht multifaktoriell konzipiert waren. Die Verwendung des HAWIE als differentielles Intelligenzdiagnostikum ist jedoch nach zahlreichen empirischen Untersuchungen (verschiedener Autoren) keineswegs unproblematisch. An eine *profilanalytische Testauswertung* werden allgemein folgende Forderungen gestellt: „Man erwartet, daß jeder Einzeltest einigermaßen reliabel ist und

daß er mit jedem anderen möglichst niedrig korreliert. Desgleichen erwartet man, daß jeder Test unabhängig vom anderen eine ausreichende logische oder empirische Validität nachweisen kann" (Lienert 1967, S. 366 f.). Diese Voraussetzungen sind beim HAWIE bzw. HAWIK nur bedingt erfüllt. So sind einmal die Testinterkorrelationen vergleichsweise hoch, zum andern ist die Subtestzuverlässigkeit (Split-half-Reliabilität) teilweise unzureichend für eine derartige profilanalytische Interpretation, wie etwa die Untersuchung von Priester & Kerekjarto (1960) zeigt. Demnach schwankt die HAWIE-Subtestreliabilität bei 20- bis 24jährigen Pbn zwischen 0.51 (schlecht) und 0.95 (brauchbar), bei 25- bis 29jährigen Pbn zwischen 0.33 (sehr schlecht) und 0.92 (noch akzeptabel) und bei 30- bis 34jährigen Pbn zwischen 0.58 (schlecht) und 0.95 (brauchbar). Die Subtestreliabilitäten des HAWIE differieren also stark, und es bestehen kaum Gründe dafür, daß die entsprechenden Verhältnisse beim HAWIK günstiger liegen (vgl. u.a. Baumeister & Bartlett 1926a u. 1962b, Kerekjarto & Schmidt 1962, Lienert & Faber 1963, Amelang & Zimmermann 1968, Zimmermann et al. 1971). Die Profilverlässlichkeit der Wechsler-Tests, die sowohl die Subtestreliabilitäten (die hoch sein sollen) als auch die Intertestkorrelationen (die möglichst niedrig sein sollen) berücksichtigt, erreicht die für eine Profilauswertung allgemein geforderte Mindestgröße von  $r_{\text{prof.}} = 0.5$  gerade noch, was aber vielfach als nicht ausreichend erachtet wird (z.B. Zimmermann et al. 1971, S. 49). Zudem dürfte diese kritische Grenze ( $r = 0.5$ ) nicht für alle Anwendungsbereiche Gültigkeit haben (Zimmermann & Kornmann 1970, Zimmermann et al. 1971). Berücksichtigt man schließlich noch die faktorenanalytischen Befunde zum HAWIE und HAWIK (s.S. 117f. oben), dann können wir abschließend festhalten: Die Wechsler-Tests messen recht gut und zuverlässig die „allgemeine“ Intelligenz sensu Spearman et al. Sie gestatten darüber hinaus einigermaßen gesichert Aussagen über die verbalen Fähigkeiten versus die Handlungsintelligenz (mehr praktische Fähigkeiten) einer Person. Spezielle Fähigkeiten wie Gedächtnis, Rechenfähigkeit u.dgl.m. sind jedoch nur bedingt mit Hilfe des HAWIE bzw. HAWIK diagnostizierbar.

**Die Haptic Intelligence Scale for Adult Blind (HISAB) von H.S. Shurrager & P.S. Shurrager (TZ)**

*Testaufbau:* Die HISAB-Serie ist in Anlehnung und analog zum WAIS-Handlungsteil speziell für Blinde von Shurrager & Shurrager in den USA entwickelt worden. Die Skala ist ein *taktile* Performancetest und soll für 16- bis 19jährige blinde oder hochgradig sehbehinderte Pbn den Handlungsteil der Wechsler-Skala (HAWIE) ersetzen, also ergänzend zum WAIS- bzw. HAWIE-Verbalteil eingesetzt werden. Der Shurrager-Test besteht aus folgenden 6 Subtests:

1. *Digit Symbol (DS)*. Dieser analog zum ZS im HAWIE entwickelte Subtest besteht aus zwei Plastiktafeln mit einfachen geometrischen Figuren, die jeweils eine bestimmte Anzahl von Punkten enthalten. Punkte und Figurkonturen sind erhaben ausgeprägt, also vom Pb haptisch-taktil wahrnehmbar. Eine der beiden Tafeln fungiert als „Schlüssel-tafel“ und trägt folgende Itemfiguren:



Die Aufgabe für den Pb besteht darin, die insgesamt 40 Items der zweiten Tafel punktzahlmäßig diesen Schlüsselitems zuzuordnen. Die Bearbeitungszeit beträgt maximal 2 Minuten.

2. *Block Design (BD)*. Das Aufgabenmaterial besteht hier aus 4 Würfeln, die je 2 glatte, 2 raue und – diagonal unterteilt – 2 halbglatte und 2 halbraue Oberflächen haben. Mit diesen Würfeln müssen bestimmte – ebenfalls taktil wahrnehmbare – Vorlagenmuster analog zum MT im HAWIE nachgelegt werden. Die Bearbeitungszeit beträgt pro Vorlage (insgesamt 7 Vorlageitems) maximal 3 Minuten.

3. *Object Assembly* (OA). Analog zum FL im HAWIE sollen hier insgesamt 4 Gegenstände (Puppe, Würfel, Hand, Ball) aus Einzelteilen zusammengelegt werden. Die Bearbeitungszeit beträgt pro Item maximal 5 Minuten.
4. *Object Completion* (OC). In Anlehnung zum BE im HAWIE werden hier 15 bekannte Gegenstände, bei denen jeweils ein wichtiger Teil weggelassen wurde, vorgegeben. Der Pb soll – ohne den Namen der betr. Gegenstände gesagt zu bekommen – jeweils das fehlende Stück benennen oder/und zeigen. Die Bearbeitungszeit beträgt maximal 1 Minute pro Item.
5. *Pattern Board* (PB). Auf einer Platte mit 5 mal 5 Löchern – in der Mitte der Platte befindet sich ein fest bzw. starr montierter Orientierungsstöpsel – werden dem Pb nacheinander 10 verschiedene Muster (Items) mit bis zu 8 Stöpseln vorgesteckt. Der Pb muß sich jedes vom Testleiter vorgesteckte Muster 1 Minute lang einprägen (abtasten) und nach Entfernung der Stöpsel dieses innerhalb der gleichen Zeit nachstecken. Die Bearbeitungszeit beträgt somit 2mal 1 Minute, also 2 Minuten pro Muster (Item).
6. *Bead Arithmetic* (BA). Hierbei handelt es sich um ein dem Abakus nachgebildetes Rechenbrett, das einen horizontalen Trennungsbalken aufweist. Dem Pb wird zunächst das System des Rechengerätes erklärt: *Unter* dem querliegenden Trennungsbalken bedeutet von rechts nach links gesehen jeder (horizontal verschiebbare) Perlenring  

1	10	100	1000	10 000
---	----	-----	------	--------

und *über* dem Trennungsbalken entsprechend  

5	50	500	5000	50 000
---	----	-----	------	--------

Der Pb tastet zunächst probeweise einige Zahlen ab, z.B. 2, 222, 333. Sodann folgen zwei Aufgabenreihen. Die erste besteht darin, daß der Testleiter nacheinander 5 Zahlen ansagt, die der Pn mit möglichst wenigen Ringen schieben soll (25, 144, 245, 612, 2089). Die zweite Aufgabenreihe besteht in der Addition von Zahlen mit Hilfe des Rechenbrettes (231+213; 552+135; 3519+120). Je nach Aufgabenschwierigkeit beträgt die Bearbeitungszeit pro Item zwischen maximal 45 Sekunden und maximal 90 Sekunden. Bei Unterschreitung der maximal angegebenen Bearbeitungszeiten wird – wie auch bei den vorausgegangenen Subtests – ein Zeitbonus gewährt.

**Erfassungsdimensionen:** Diesbezügliche Hinweise fehlen im Testmanual fast ganz. Mutatis mutandis dürften jedoch hier die schon beim HAWIE angeführten Leistungsdimensionen einschlägig sein. Siehe auch die Ausführungen zur Validität der HISAB unten.

**Testgütekriterien:** Instruktion und Auswertungskriterien sind genau fixiert; die *Objektivität* der HISAB ist somit gewährleistet.

Zur *Reliabilität* der HISAB machen die Testautoren folgende Angaben:

HISAB-Subtests:	<i>Split-half</i> -Reliab.	<i>Retest</i> -Reliab. (nach 6 Mon.)
1. Digit Symbol	0.93	0.77
2. Block Design	0.93	0.77
3. Object Assembly	0.79	0.70
4. Object Completion	0.85	0.76
5. Pattern Board	0.94	0.81
6. Bead Arithmetic	0.94	0.75

Für den Gesamttest beträgt die Halbierungszuverlässigkeit 0.91. Damit genügt die HISAB-Reliabilität den üblichen Anforderungskriterien.

Die *Validität* des Tests wurde durch Korrelation der HISAB mit dem WAIS-Verbalteil ermittelt: Mit  $r = 0.65$  besteht immerhin ein mittlerer Zusammenhang zwischen *verbalem* Teil der WAIS und Gesamtergebnis im HISAB. Die Subtests der HISAB sind mit den Subtests der WAIS zwischen 0.31 und 0.61 korreliert, während die HISAB-Intertestkorrelationen sich zwischen 0.53 und 0.82 bewegen, also vergleichsweise hoch ausfallen. Dies legt den Gedanken nahe, daß sämtliche 6 Subtests – zum überwiegenden Teil – an einem gemeinsamen (g-)Faktor beteiligt sind. „It may be that ability to solve concrete problems

by perception and manipulation of objects and relationships without the aid of sight will always involve to a large extent a common factor" (Shurrager & Shurrager 1964, S. 22). Ferner wird darauf hingewiesen, daß die Interkorrelationen von Tests, die motorische Geschicklichkeit verlangen, bei *blinden* Pbn höher ausfielen als die Interkorrelationen der gleichen Tests, die auf Ergebnissen *sehender* Pbn basieren. Analoge Tendenzen zeigen sich bei den Intertestkorrelationen des WAIS-Verbalteils.

Solange keine deutschsprachigen Eichwerte (*Normen*) zur Verfügung stehen, ist man auf die bei 700 amerikanischen Blinden im Alter von 16 bis 19 Jahren erhobenen Normdaten (Abweichungs-IQ) angewiesen.

**Anwendungsbereiche:** Der Test eignet sich zur Erfassung der allgemeinen Intelligenz bzw. Handlungsintelligenz 16- bis 19jähriger blinder oder hochgradig sehbehinderter Jugendlicher. Die HISAB sollte jedoch stets zusammen mit dem Verbalteil des HAWIE durchgeführt werden.

**Kritische Würdigung der HISAB:** Das Testverfahren ist eine wertvolle Ergänzung zum HAWIE-Verbalteil in der Anwendung bei blinden oder/und hochgradig sehbehinderten Personen im Alter von 16 bis 19+ Jahren. Das HISAB-Diagnostikum ist objektiv und – zumindest im Gesamtergebnis – reliabel und valide. Eine Normenkontrolle, u.U. sogar eine Aufgabenanalyse, im Hinblick auf deutschsprachige Verhältnisse erscheint jedoch dringend geboten. Für eine (eventuelle) profilanalytische Auswertung des HISAB-Diagnostikums gelten die gleichen Kautelen, wie sie schon im Zusammenhang der HAWIE-Erörterung erhoben wurden. Der Shurrager-Test ist in erster Linie ein Diagnostikum zur Erfassung der allgemeinen Begabung resp. einer (mehr praktischen) Handlungsintelligenz, wie sie beispielsweise im Performanceteil des Wechsler-Tests zum Ausdruck kommt.

**Sprachfreie Intelligenzuntersuchung für Hörende und Taubstumme oder Snijders-Oomen Nicht-verbale Intelligenztestreihe (SON) von J.T. Snijders & N. Snijders-Oomen (TZ)**

Der SON-Test gehört zu den Handlungsskalen und kann darüber hinaus „sprachfrei“ instruiert und durchgeführt werden. Inwieweit eine wirklich sprachfreie bzw. (besser:) sprachunabhängige Testuntersuchung innerhalb unseres Kulturraumes – auch bei nicht sprechenden Populationsangehörigen, z.B. Taubstummen – überhaupt möglich (und notwendig) ist, wurde u.a. von Neumann (1968) ausführlich diskutiert. Immerhin kann das SON-Diagnostikum ohne Informationsverluste bei Hör- und Sprachbehinderten durchgeführt werden. Es ist außerdem bislang das einzige Intelligenzmeßverfahren im deutschsprachigen Raum, das spezielle Testnormen für hörsprachgeschädigte Kinder und Jugendliche anbietet. Der SON besteht aus zweimal vier parallelen Subtests, die zwar durchnummeriert sind, jedoch eigentlich als *parallele* Testreihen konzipiert wurden (Testreihe P versus Q). Man kann also beide Testreihen zusammen (8 Subtests) oder eine der beiden Testreihen (gewöhnlich Reihe P) durchführen. Somit ergibt sich folgende Zusammensetzung des SON:

1. **Mosaik:** Mosaik A und B, Würfelfiguren (P)
2. **Zeichnen:** Nachzeichnen, Figurergänzung (Q)
3. **Kombination:** Zusammensetzspiel, Bilderreihen A und B (P)
4. **Ergänzung:** Hälften, Zusammengehörige Karten, Bilderergänzung (Q)
5. **Analogien:** Reihen fortsetzen, Bilderanalogien, Figurenanalogien (P)
6. **Sortieren:** Plättchen sortieren, Karten sortieren (Q)
7. **Gedächtnis für Karten:** Gedächtnis für Karten A und B (P)
8. **Knox Würfel:** Merkfähigkeitsprüfung (Q)

Die Aufgabenzusammenstellung im SON erfolgte also in enger Anlehnung an die Handlungsskalen gängiger (Binet-)Intelligenztests.

**Erfassungsdimensionen:** Die Testautoren ordnen die 8 Subtests 4 Leistungsdimensionen oder Intelligenzfunktionen zu, nämlich I. *Form* (Subtests 1+2), II. *Anschaulicher Zusammen-*



*hang* (Subtests 3+4), III. *Abstraktion* (Subtests 5+6) und IV. *Unmittelbares Gedächtnis* (Subtests 7+8). „Je zwei der acht Subtests sind nach psychologischen Gesichtspunkten zu *vier Testgruppen* zusammengefaßt: Form, anschaulicher Zusammenhang, Abstraktion und unmittelbares Gedächtnis. Die Bedeutung dieser Namen wird aus den nachfolgenden (hier nicht wiedergegebenen; d. Verf.) Beschreibungen deutlich werden. Diese vier Aspekte der Intelligenz zeigten sich bei den Vorbereitungen der ersten Form der Testreihe (bereits 1939-42; K.H.) als wichtig für die Intelligenztypisierung taubstummer Kinder. Bei der Revision der Testreihe wurde diese Einteilung übernommen. Die faktorenanalytische Verarbeitung der Eichungsunterlagen hat ihre Bedeutung nur teilweise bestätigen können. Trotzdem wurde sie vorläufig beibehalten, auch darum, weil eine *Andeutung* klinischer Brauchbarkeit vorliegt“ (Snijders & Snijders-Oomen 1958b, S. 8). Letztere Begründung dürfte allerdings kaum als Argument für eine differentialdiagnostische (profilanalytische) Auswertung des SON ausreichen, der somit den allgemeinen Intelligenztests zugerechnet werden muß.

*Testgütekriterien:* Der SON ist verhältnismäßig sorgfältig standardisiert worden. Die Eichpopulation umfaßte ausschließlich holländische Kinder im Alter von 3 bis 16 Jahren, und zwar 1400 normalhörende und 1054 taubstumm Pbn. Kleinere Kontrollerhebungen liegen u.a. aus Belgien und Deutschland vor und erlauben trotz gewisser Repräsentativitätsmängel der Samples den Schluß, daß die Benutzung der an holländischen Kindern und Jugendlichen gewonnenen *Testnormen* im deutschsprachigen Raum – nicht zuletzt im Hinblick auf die „sprachfreie“ Intelligenzuntersuchung – kein nennenswertes Risiko birgt. Als SON-Normen werden einmal durch Rohwerttransformation gewonnene *normalisierte Standardwerte* (StW), zum andern (Abweichungs-) *Intelligenzquotienten* (IQ) angeboten. Während die IQ-Skala in Anlehnung an Wechsler mit  $M_{IQ} = 100$  und  $s_{IQ} = 15$  definiert ist, wird die Standardwertskala hier – anders als etwa im Intelligenzstrukturtest (IST) von Amthauer – mit  $M_{StW} = 25$  und  $s_{StW} = 5$  angegeben. Ferner werden Intelligenzalter (IA) und Subtestalter (SA) resp. Entwicklungsalter verwendet. Die etwas antiquiert wirkende Normierungsprozedur beschreiben die Testautoren folgendermaßen. „Für die Berechnung des IQ ist das Intelligenzalter (IA) nicht erforderlich. Wir ziehen daraus nicht, wie Wechsler, die Konsequenz, daß das IA weiterhin keine Bedeutung hat. Es hat einen selbständigen Sinn als Ausdruck für die Entwicklungslage des Kindes. Während Standardwert und IQ sich auf die Stellung der Vp innerhalb ihrer eigenen Altersgruppe beziehen, hat das IA Bezug auf ihre Stellung in der Entwicklungsreihe. Die Berechnung verläuft in zwei Schritten: a) In Tab. 58-59 (Seite 144) findet man für jeden möglichen Rohwert von jedem Subtest das Entwicklungsalter, d.h. das Alter, in dem 50 % der Vpn diesen Wert erzielen. Dieses benennen wir *Subtestalter* (SA). b) Aus den Subtestalterswerten berechnet man zusammenfassend einen mittleren Wert: das *Intelligenzalter* (IA). Für die höheren Rohwerte und die höheren Altersstufen hat der Begriff des IA keinen Sinn mehr. Darum wird kein IA berechnet, das höher wäre als 13“ (loc. cit., S. 13). Siehe auch nachstehend abgebildetes Auswertungsmuster zum SON. Die im Testmanual beigelegten Normentabellen (zur Umsetzung der Rohwerte in StW bzw. IA sowie Transformation der StW-Summen in IQ) reichen von 4;11 bis 15;6 J., wobei die Klassenintervalle von zunächst 2 Monaten mit zunehmendem Alter ansteigen und zuletzt 12 Monate (1 Klassenintervall) umfassen. Für Hörende und Gehörlose werden jeweils getrennte SON-Normen angeführt.

Ebenfalls getrennt für Hörende und Taubstumm wurden die *Reliabilität* des SON sowie die Meßfehler berechnet. Demnach liegt die Halbierungsreliabilität für die Gesamtskala zwischen 0.91 und 0.95, für die verschiedenen Subtests bzw. Altersstufen jedoch nur zwischen 0.45 und 0.93. Mit 78 gehörlosen Kindern wurde im Abstand von drei Jahren 1956 eine Retestung durchgeführt. Die entsprechenden (Retest-)Reliabilitäten betragen auf den IQ bezogen für die Gesamttestreihe 0.80, für die Reihe P 0.81 und für die Reihe Q (nur) 0.63. Reihe P erwies sich somit bedeutend zeitstabiler als Reihe Q.

Der *Standardmeßfehler* beträgt nach den Berechnungen der Testautoren im Mittel 2,5 StW für die 8 Subtests und gilt sowohl für Hörende als auch für Taubstumm. Allerdings schwanken die Meßfehler von Subtest zu Subtest resp. zwischen den verschiedenen Altersgruppen

# SNIJDRS-OOMEN NICHT-VERBALE INTELLIGENZ-TESTREIHE.

Stand. Wert	Mosaik	Zeichnen	Kom- bination	Ergänzung	Analogien	Sortieren	Gedächtnis Karten	Knox Würfel
45								
44								
43								
42								
41								
40								
39								
38								
37								
36								
35								
34								
33								
32								
31			X					
30								
29								
28								
27								
26								
25		X		X				
24								
23	X							
22								
21						X		X
20					X			
19								
18								
17							X	
16								
15								
14								
13								
12								
11								
10								
9								
8								
7								
6								
5								

← Achte auf die geänderte Reihenfolge.

Alter.....12;2.....

Summe St. W.  
der abgenommenen Subtests. . . . .183

Anzahl abgen. Subtests . . . . .8

I.Q.: .....90.....

S.A. der Subtests vom höchsten bis zum  
niedrigsten

.....>15.....

.....12.....

.....12.....

.....10,5.....

.....9.....

.....9.....

.....6,5.....

I.A. ....10,5.....

Zusammenfassung:

Muster zur Auswertung des SON (Snijders-Oomen Nicht-verbale  
Intelligenz-Testreihe)

teilweise beträchtlich und reichen von 1,32 StW (Mosaik bei 5 – 5 1/2-jährigen Taubstummen) bis 4,0 StW (Knox Blöcke bei 10 – 11-jährigen Hörenden). Entsprechend werden für die drei Altersgruppen I (5 – 5 1/2 J.), II (10 u. 11 J.) und III (15 J.) folgende, jetzt auf *Intelligenzquotienten* (IQ) bezogene Standardmeßfehler für Hörende (H) versus Taube (T) mitgeteilt:

	I		II		III	
<i>P-Reihe</i> (4 Subtests):	4 (H)	4 (T)	5 (H)	5 (T)	5 (H)	4 (T)
<i>Q-Reihe</i> (4 Subtests):	5 (H)	5 (T)	7 (H)	6 (T)	7 (H)	5 (T)
<i>Gesamttest</i> (8 Subtests):	3 (H)	4 (T)	4 (H)	5 (T)	5 (H)	4 (T)

Für die *gesamte* SON-Skala beträgt somit der Standardmeßfehler durchschnittlich + 4 IQ-Punkte. Auch fällt hier wieder die bessere Zuverlässigkeit der P-Reihe (gegenüber der Q-Reihe) auf, wenngleich diese in ihrer Aussagezuverlässigkeit – erwartungsgemäß – unter der Meßgenauigkeit der Gesamtskala bleibt. Ferner sinkt die Meßgenauigkeit der SON mit ansteigendem Alter der Pbn.

Die *Validität* des SON, ermittelt durch Korrelation mit dem Lehrerurteil, liegt (je nach Altersgruppe bzw. Schulstufe) bei Hörenden und Gehörlosen zwischen 0.38 und 0.57, wobei allerdings die Irreliabilität des Lehrerurteils als wertmindernder Faktor hier in Rechnung zu stellen ist. Bei vollstandardisierten Intelligenztests als Außenkriterium sind durchaus höhere Validitätskoeffizienten zu erwarten, nach unseren Berechnungen (vgl. Heller 1967b, S. 227) Werte um 0.80 in der Relation SON-HAWIK-Handlungsteil sowie Werte zwischen 0.53 und 0.89 in der Relation SON-PMT bei hörgeschädigten Kindern und Jugendlichen.

*Anwendungsbereiche:* Das SON-Diagnostikum eignet sich zwar grundsätzlich für die (allgemeine) Intelligenzmessung Tauber *und* Hörender – was z.B. bei Gruppenvergleichen interessant ist –, in praxi wird der Snijders-Oomen-Test jedoch fast ausschließlich bei *hör-* und/oder *sprachgeschädigten* Kindern und Jugendlichen im Alter von 3 bis 16 Jahren angewandt. Dabei erweist sich die größere Treffsicherheit des Diagnostikums offensichtlich in seiner Anwendung bei jüngeren Kindern (siehe die Ergebnisse zur Reliabilitäts- und Meßfehlerkontrolle

*Kritik:* Der Snijders-Oomen-Test ist ein Test zur Erfassung der *allgemeinen Intelligenz*. Für diesen Zweck ist er valide und zuverlässig. Sein Vorzug liegt in der Möglichkeit der nonverbalen Administration, was ihn besonders in der Hör-/Sprachgeschädigtendiagnostik interessant macht. Andererseits haften dem Test unverkennbare Mängel an. Er ist sowohl in der Durchführung als auch in der Auswertung zu aufwendig und auch unökonomisch; der Arbeitsaufwand steht in keinem rechten Verhältnis zur Effektivität des Tests. Noch stärker als beim HAWIK treffen hier die Bedenken in bezug auf eine differentialdiagnostische Auswertung (Profilanalyse der Subtests) zu. Trotz formal gegebener Möglichkeit unabhängiger Subtestauswertung (durch StW-Normierung) sprechen sowohl die verhältnismäßig niedrigen und stark schwankenden Subtestreliabilitäten als auch vermutlich sehr hohe Testinterkorrelationen versus die fehlende empirische Verifikation der von den Testautoren postulierten vier Erfassungsdimensionen übereinstimmend *gegen* eine profilanalytische Auswertung des SON-Diagnostikums. Zur Bestimmung der *allgemeinen Intelligenz* ist der SON, besonders in der Gesamtform (8 Subtests) resp. in der Form P, weniger in der Form Q, recht gut geeignet. Allerdings stehen uns heute rationellere Intelligenzmeßverfahren, auch in der Hör-/Sprachgeschädigtendiagnostik, zur Verfügung (vgl. Kap. III, 2b unten). Für die jüngeren Kinder im Vorschul- und Grundschulalter, die Hör- und/oder Sprachbehinderungen aufweisen, wird der Snijders-Oomen-Test nach wie vor zum Repertoire intelligenzdiagnostischer Verfahren gehören.

## b) Gruppentests

**Aufgaben zum Nachdenken** (AzN 4+) von E. Hylla & B. Kraak, Neubearbeitung von H. Horn, E. Schwarz und U. Raatz (Beltz Verlag)

**Testaufbau:** Der AzN 4+ ist als „Begabungstest für den Übergang auf weiterführende Schulen“ konzipiert und hat sich als solcher auch bewährt. Er besteht aus 5 Subtests, nämlich Rechnen (RE), Analogien (AN), Zahlenreihen (ZR), Satzergänzung (SE) und Instruktionsverständnis (IV).

**Erfassungsdimensionen:** Faktorenanalytische Untersuchungen zum AzN sind nicht bekannt. Der Test mißt – nach allen vorliegenden Erfahrungen – die Begabung für weiterführende Bildungseinrichtungen, also wesentliche Merkmale der Gymnasial- und Realschuleignung (bzw. Hauptschuleignung). Die formal-*inhaltliche* Struktur des Tests deutet (eher) ein *Mixtum compositum* aus Intelligenz- und Lernleistungstest (als ein „reines“ Intelligenzdiagnostikum) an. Aus dem Aufgabencharakter der jeweiligen Subtestitems lassen sich im einzelnen psychologisch folgende Dimensionen – akzentuierend – ableiten:

- RE    Rechenfähigkeit; numerischer Intelligenz-Leistungsfaktor; Abstraktions- und Denkfähigkeit i.w.S.
- AN    Sprachlogische und kombinatorische Fähigkeiten im Umgang mit verbalem Aufgabenmaterial
- ZR    Schlußfolgerndes Denken nonverbaler Art; numerischer Faktor; g-Faktor sensu Spearman et al.
- SE    Verbaler Faktor; Verbal reasoning; Wortsinnverständnis
- IV    Komplexer Sprachfaktor; Erfassen komplizierter Zusammenhänge und Instruktionsverständnis; Denkfähigkeit(en)

**Testgütekriterien:** Der AzN-Test ist ein *objektives* Verfahren, d.h. sowohl die Standardbedingungen der Testdurchführung als auch die Auswerterreliabilität sind optimal, sofern sich die Testleiter (Lehrer) genau an die im Testmanual fixierten Vorschriften zur Administration und Auswertung des Tests halten.

Die *Reliabilität* des AzN ist gut. Die Split-half-Reliabilitätskoeffizienten liegen (nach Spearman-Brown aufgewertet) bei 0.92 (Testform A) und 0.88 (Paralleltestform B). Die Konsistenzanalyse (nach der Kuder-Richardson-Formel) erbrachte sogar Reliabilitätskoeffizienten von 0.95 (Form A) und 0.93 (Form B). Die Paralleltestzuverlässigkeit liegt allerdings nur bei  $r = 0.85$ . Die Retestmethode wäre angesichts der für einen Begabungstest relativ hohen Lernabhängigkeit des AzN-Ergebnisses fragwürdig und wurde wohl deshalb auch nicht angewendet. Der *Standardmeßfehler* beträgt 3,4 Rohpunkte.

Zur *Validität* des AzN 4+ werden im Testmanual Bewährungskontrollen aus verschiedenen Untersuchungen zur Schuleignungsermittlung angeführt. Die höchsten (biserialen) Validitätskoeffizienten liegen nach einer Untersuchung von Gebauer aus dem Jahre 1964 bei 0.62. Dieses Ergebnis besagt, daß der AzN unter den seinerzeit verwendeten Untersuchungsmethoden (außer AzN wurden noch das Lehrerurteil, der Kretschmer-Höhn-Test, der Frankfurter Wortschatztest, der sog. Stuttgarter Lückentest und der Pauli-Test in den Methodenvergleich einbezogen) zwar am besten – am zweitbesten das Lehrerurteil mit Koeffizienten zwischen 0.4 und 0.5 – abschnitt und eine Kombination von AzN *und* Lehrerurteil erwartungsgemäß noch höhere (multiple) Validitätskoeffizienten ergeben würde, es ist jedoch andererseits als *optimales* und damit nicht repräsentatives Gültigkeitskriterium hier zu werten. Nach unseren – im Rahmen der baden-württembergischen Schullaufbahnberatung initiierten – wesentlich umfangreicheren Bewährungskontrollen verschiedener Prädiktoren (vgl. u.a. Heller 1970a und 1972d, Allinger & Heller 1972) fallen entsprechende Validitätskoeffizienten insgesamt, d.h. in bezug auf sämtliche verfügbaren Auslesemethoden einschließlich des AzN, *durchschnittlich* niedriger aus (siehe noch Tent 1969). Die Interkorrelationen der Subtests liegen mit Werten zwischen 0.49 und 0.63 für ein (im Falle des AzN ursprünglich nicht vorgesehenes) Profilverfahren zu hoch; die einzelnen Untertests korrelieren mit dem Gesamttestergebnis sogar zwischen 0.78 und 0.86. Diese Resultate deuten somit unter dem Gesichtspunkt der Aufgabenstruktur eine recht gute Homogenität des AzN an, sie verbieten aber andererseits praktisch eine unabhängige Subtestinterpretation. Der AzN dient also in erster Linie der Erfassung mehr allgemeiner

Begabungsqualitäten der Gymnasial- und Realschuleignung versus indiziert das allgemeine intellektuelle (Lern-)Leistungsniveau im 4. Grundschuljahr.

Für den AzN wurden 1967 neue *Normen* berechnet. Somit liegen nunmehr 1) Altersnormen (in PR) für 9;0 bis 11;0 J. vor, 2) Klassennormen (in PR) für 4. Grundschulklassen vollausgebaute versus nicht-vollausgebaute Schulen und 3) Teiltestnormen (Stabdiagramme der Median- und Quartilwerte für die 5 Subtests des AzN). Demnach liegen die *Medianwerte* (in Klammern  $Q_1 = 25$ . PR und  $Q_3 = 75$ . PR) bei folgenden Rohwerten (RW): in RE bei 5,8 RW (3,9 u. 7,7), in AN bei 13,6 RW (9,4 u. 16,4), in ZR bei 10,3 RW (7,7 u. 12,8), in SE bei 7,4 RW (5,0 u. 9,4) sowie in IV bei 7,2 RW (4,5 u. 9,3). Eine im Testmanual beigelegte Transformationstabelle erlaubt die Umwandlung der PR-Einheiten in T-Werte oder IQ; weitere Umwandlungen sind mit Hilfe unserer im Appendix als Tab. VIII aufgeführten Transformationstabelle von Testnormen (s.S. 245) möglich.

**Anwendungsbereiche:** Der AzN eignet sich vor allem zur Schuleignungsermittlung resp. Begabungsdifferenzierung im 4. Grundschuljahr. Im 5. Schuljahr ist er – entgegen der Etikettierung (4+) – nur mit Vorbehalten zu verwenden, da hierfür keine ausreichenden Normen zur Verfügung stehen und wohl auch keine angemessene Schwierigkeit der Testitems mehr gegeben ist. Wegen der relativ hohen Testempfindlichkeit gegenüber früheren Lernerfahrungen resp. gegenüber sozio-kulturellen Umwelteinflüssen überhaupt sollte das AzN-Diagnostikum immer in Verbindung mit anderen (auch nonverbalen) Intelligenztests durchgeführt werden.

**Kritik:** Der AzN gehört zu den klassischen „Auslese“-Verfahren. Berücksichtigt man, daß in die AzN-Testleistung mehr oder weniger unkontrollierbare Milieueinflüsse eingehen, so kann man den AzN (nur) als Zusatzverfahren im Rahmen der Schuleignungsermittlung resp. Begabungsdifferenzierung auf der 4. Grundschulstufe empfehlen. Als solches aber hat sich der AzN vielfach bewährt und stellt nicht nur ein brauchbares, sondern auch ein sehr nützliches Instrument der Begabungsdifferenzierung dar. Eine – formaliter (Unabhängigkeit der Subtestauswertung) mögliche – Profilanalyse des AzN stößt dagegen auf erhebliche Schwierigkeiten, nicht nur auf Interpretationsebene. Angesichts der erheblichen Redundanz einzelner Testteile (vgl. oben aufgeführte Erfassungsdimensionen) und ihrer fehlenden faktorenanalytischen Verifikation im Sinne eines differentiellen Begabungstests sowie unter Berücksichtigung der hohen Interkorrelationen der Testteile untereinander und in bezug auf den Gesamttest lassen die mitgeteilten Parameter der Eichpopulation (Median- und Quartilwerte in den Subtestleistungen) allenfalls eine sehr grobe oder besser: akzentuierende Interpretation der AzN-Subtestdimensionen zu.

#### **Culture Free or Culture Fair Intelligence Tests (CFT) von R.B. Cattell (TZ)**

Von R. Cattell stammen mehrere sog. kulturunabhängige Intelligenztests, so der „IPAT Culture Free Intelligence Test, Scale 1“ für Kinder im Alter von 4 bis 8 Jahren, der „IPAT Culture Free Intelligence Test, Scale 2“ für Kinder im Alter von 8 bis 13 Jahren, der „IPAT Culture Free Test, Scale 3“ für Jugendliche und Erwachsene. Letztere Form ist inzwischen, für deutsche Verhältnisse adaptiert und herausgegeben von R. Weiß, im G. Westermann Verlag erschienen (vgl. Cattell & Weiß 1970). Hier soll nur Skala 2 besprochen werden (vgl. Cattell 1962 sowie Cattell & Cattell 1957).

**Testaufbau:** Die CFT-Skala 2 besteht aus zwei Testformen (Form A und B), die zwar häufig als Parallelformen benutzt werden, ursprünglich jedoch vom Testautor nicht als solche, sondern als ergänzende – sukzessiv darzubietende – Skalenhälften vorgesehen waren. Jede Form besteht aus 4 Subtests mit einmal 14, zweimal 12 und einmal 8 Aufgaben. Die Lösung der Aufgaben erfolgt sprachfrei oder nonverbal, nicht jedoch die Instruktion der Aufgaben. Das Aufgabenmaterial besteht im wesentlichen aus Matrizen, wobei figurale Serien- und Klassifikationsitems sowie geometrische Form-Analogien überwiegen. Zur Veranschaulichung seien hier einige *Übungsbeispiele* abgebildet.

Die Lösung der einzelnen Aufgaben wird auf einem separaten Antwortbogen eingetragen; wie bei fast allen modernen Papier-Bleistift-Verfahren (Paper and Pencil Tests) wurde auch

hier die *Mehrfachwahl-Antwortform* gewählt: Der Pb streicht oder kreuzt die unter den angebotenen Antworten jeweils (einzig) richtige Lösung an. Unter anderem wird dadurch die Auswertung vereinfacht und maximal objektiviert. Ferner ist bei diesem Verfahren grundsätzlich eine elektronische Datenauswertung (Belegauswertung) möglich – und bei umfangreicheren Untersuchungen notwendig oder gar unerlässlich.

1	2	3	4	5	ANSWERS
					<input type="checkbox"/> 4
1	2	3	4	5	ANSWERS
					<input type="checkbox"/> 3
					<input type="checkbox"/> 1
					<input type="checkbox"/> 3
1	2	3	4	5	
					<input type="checkbox"/>

**Erfassungsdimensionen:** Der CFT wird im allgemeinen als guter Indikator der allgemeinen Intelligenz (g-Faktor sensu Spearman et al.) angesehen. Als solcher hat er sich auch im Rahmen der baden-württembergischen Bildungsberatung bewährt, und zwar besonders dort, wo sozio-kulturelle Einflüsse der Intelligenzentwicklung isoliert werden sollten. Trotzdem: Der CFT ist *kein* kultur-unabhängiger Test i.e.S.

Aufgrund eigener faktorenanalytischer Untersuchungen zum CFT 3 kommt Weiß auf drei die CFT-Leistung wesentlich konstituierende Dimensionen, die er **zusammenfassend** – sensu R.B. Cattell – „Grundintelligenz“ nennt: 1) Reasoning, 2) Beziehungsstiftendes Denken, 3) Interferenzneigung/funktionelle Störbarkeit. Diese drei Faktoren umfassen zusammen rd. 2/3 der Gesamtvarianz des CFT (Cattell & Weiß 1971, Weiß 1969).

**Testgütekriterien:** Hierüber liegen, vor allem im deutschsprachigen Raum, sehr widersprüchliche Angaben vor. In der Literatur werden *Reliabilitätskoeffizienten* zwischen 0.70 und 0.92 erwähnt, auch wird die *Objektivität* bzw. Ökonomie des CFT positiv hervorgehoben, wohingegen bezüglich der *Validität* vielfach Zweifel angemeldet werden (vgl. u.a. Anastasi 1968, Wewetzer 1964, Pawlik 1970). Unsere Erfahrungen bei der Übertrittsauslese bestätigen die Objektivität und Ökonomie des CFT, wohingegen einzelne Reliabilitätskontrollen (anhand nichtrepräsentativer Samples des 4. Schuljahres) stärker als erwünscht streuen, jedoch im oben skizzierten Rahmen liegen. Mit dem IST (Gesamtstandardwert) korreliert der CFT zwischen 0.6 und 0.7 mit dem PSB (Gesamtleistung) mit 0.4 bis 0.65 (Cattell & Weiß 1971, S. 38). Unter den Schulleistungsfächern ist Mathematik am höchsten mit der CFT-Gesamtleistung korreliert, nämlich mit 0.53 (Gymnasialzensuren) und 0.58 (Realschulzensuren). Weitere Kontrollergebnisse sind im Zusammenhang der Publikation des CFT 2, die ebenfalls R. Weiß besorgt, zu erwarten. Siehe auch Aurin et al. 1968.

Als *Normen* werden in der Originalform IQ (Altersnormen) angeboten. Die bei einer größeren Erhebung im 4. Grundschuljahr 1967 in Baden-Württemberg gewonnenen Normdaten erbrachten folgende Resultate (jeweils auf das CFT-Gesamtergebnis bezogen):

Altersgruppe 9;0 – 9;5	M = 15,5 RW;	s = 5,2 RW	(N = 287)
Altersgruppe 9;6 – 9;11	M = 17,0 RW;	s = 6,3 RW	(N = 475)
Altersgruppe 10;0 – 10;11	M = 17,0 RW;	s = 5,0 RW	(N = 293)
4. Grundschulklasse (total)	M = 16,5 RW;	s = 5,9 RW	(N = 1055)

**Anwendungsbereiche:** Der CFT empfiehlt sich überall dort, wo man nonverbale Intelligenzfaktoren, vorab Reasoning und „beziehungsstiftendes Denken“, oder Begabungspotenzen (z.B. zur Aufhellung sog. Underachievement-Probleme) diagnostizieren will. Durch relativ geringfügige Modifikationen an der Instruktion dürfte die Anwendung des CFT auch bei Hör- und/oder Sprachgeschädigten möglich sein und als Indikator der „allgemeinen intellektuellen Leistungsfähigkeit“ interessant werden. Im Rahmen der Schul- und Studieneignungs-ermittlung spielt der CFT dagegen wohl nur teilweise eine wichtige Rolle, etwa bei der Erfassung vorhandener, sozio-kulturell jedoch mehr oder weniger vernachlässigter Begabungspotenzen.

**Kritik:** Die CFT-Skalen sind *nicht* – wie der Name beansprucht – kulturunabhängige Intelligenzmeßverfahren, obgleich ihnen eine gewisse Universalität (was die geographische Verbreitung, nicht die Anwendungshäufigkeit versus Dimensionalität erfaßter Intelligenzfaktoren betrifft) nicht abgesprochen werden kann. Kulturunabhängige Intelligenztests wird es im strengen Sinne nie geben (vgl. auch unsere Ausführungen im theoretischen Teil, S. 35 ff.). Jedoch sind die Aufgabeninhalte des CFT praktisch *nonverbaler* Art.

Der CFT ist ein ökonomisches und objektives Verfahren zur Erfassung der allgemeinen Intelligenz im Sinne der Zweifaktorentheorie, wenngleich er offensichtlich andere Aspekte der Allgemeinbegabung als beispielsweise der Wechsler-Test mißt. Hinsichtlich der Reliabilität und vor allem der Validität bleiben noch manche Wünsche offen, sieht man einmal von den vielleicht zu positiv bewerteten Untersuchungsergebnissen von Weiß (1969) und Cattell & Weiß (1971) ab.

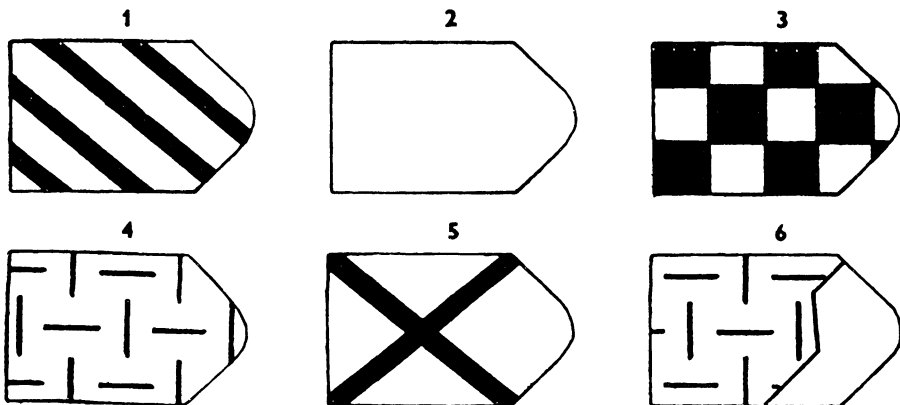
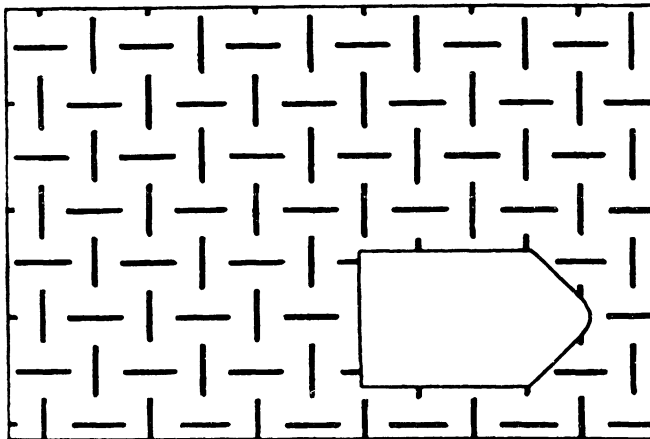
### **Der Progressive Matrizen-Test (PMT) oder The Progressive Matrices von J.C. Raven (TZ)**

Die erste Auflage des PMT erschien 1938. Inzwischen sind zahlreiche Revisionen sowie zwei Spezialformen für Kleinkinder und überdurchschnittlich begabte Erwachsene veröffentlicht worden, so daß nunmehr drei Formen existieren: eine farbige Testform (*Coloured Progressive Matrices*), erstmals 1947 publiziert, mit drei Aufgabenreihen (A, AB, B) für 5 bis 11jährige Kinder sowie schwachsinnige Erwachsene; die Standardform (*Standard Progressive Matrices*) mit 5 Reihen (A, B, C, D, E) zu je 12 Aufgaben für Kinder ab dem 6. bzw. 8. Lebensjahr sowie Jugendliche und Erwachsene bis zum 65. Lebensjahr; eine Spezialform für Hochbegabte (*Advanced Progressive Matrices*) bzw. höher begabte Erwachsene mit zwei Sets (I und II).

**Testaufbau:** Es wird eine Tafel mit einem bestimmten Muster jeweils vorgegeben; dieses ist unvollständig, d.h. ein Teil des Musters ist weggelassen. Der Pb soll nun aus den vorgegebenen 6 oder 8 Teilstücken das jeweils passende herausfinden und die Lösungsnummer auf einem separaten Antwortbogen notieren. Siehe nachstehendes Übungsbeispiel zum non-verbalen PMT von Raven.

Die Aufgaben sind dem Schwierigkeitsgrad nach aufsteigend – sowohl innerhalb der einzelnen Sets als auch zwischen den Serien „progressiv“ – angeordnet. Allerdings ließ sich diese Progression empirisch nicht ganz verifizieren. Der PMT ist ein reiner *Power*test, d.h. er wird ohne zeitliche Limitierung bearbeitet (vgl. S. 65, Fußn. 40). In der englischen Originalausgabe (Testmappe) ist noch der Mill-Hill-Vocabulary-Test (als verbaler Zusatztest von Raven selbst empfohlen) beigelegt, der jedoch außerhalb des englischen Sprachraumes kaum Beachtung findet.

**Erfassungsdimensionen:** Nach Angaben des Autors (Testmanual) soll der PMT „clear thinking“ prüfen; auch war der PMT ursprünglich eher als Entwicklungstest denn als Intelligenztest konzipiert<sup>50</sup>. Faktoranalytische Untersuchungen erbrachten jedoch den Nachweis, daß der PMT sehr gut die *allgemeine Intelligenz* (g-Faktor sensu Spearman et al.) erfaßt. Darüberhinaus sind vor allem die Faktoren *induktives Denken* und *Raumvorstellung* an der PMT-Leistung beteiligt. Für die Interpretation der PMT-Ergebnisse im Anwendungsfalle bedeutet dies, daß – neben der Allgemeinbefähigung – folgende Intelligenzfunktionen im PMT zum Ausdruck kommen: Unterscheidungsgenauigkeit (accuracy of discrimination) – vorab bei den einfacheren Aufgaben; Analogieschlußdenken, Regelerkennen bzw. Abstrahieren von Prinzipien und deren (richtige) Anwendung, induktives Denken sowie Raumvorstellungskomponenten – vorab bei den komplexeren Aufgaben.



<sup>50</sup> Persönliche Information des Testautors anlässlich eines Deutschlандаufenthaltes im Jahre 1966.



**Testgütekriterien:** Im allgemeinen, d.h. bei korrekter Administration resp. Testdurchführung, ist die *Objektivität* des PMT gewährleistet. Bei nonverbaler Instruktion, z.B. bei Gehörlosen oder Aphasikern, muß besonderer Wert auf das Instruktionsverständnis seitens des Pb gelegt werden. Durch zusätzliche Veranschaulichungshilfen (etwa Vorlage der Aufgabenbeispiele als *Formbretter* oder unterstützende Ausdruckshilfen, z.B. Gebärden) lassen sich auftretende Verständnisschwierigkeiten erfahrungsgemäß relativ leicht ausräumen.

Zur *Reliabilität* des PMT finden sich in der Literatur unterschiedliche Angaben. Die berichteten Zuverlässigkeitskoeffizienten schwanken zwischen 0.9 und 0.7, was gute bis noch befriedigende Testreliabilität bedeutet. Bei deutschen Taubstummen fanden wir (Heller 1967b, S. 227) Stabilitätskoeffizienten von 0.98 (Retestung nach 4 Wochen), 0.85 (Retestung nach 9 Monaten) und 0.72 (Retestung nach 2 Jahren). Bangen (1965, S.41) berechnete bei 9 – 10jährigen Taubstummen eine Halbierungszuverlässigkeit von  $r = 0.90$ . Somit ist die Zuverlässigkeit des PMT in der Anwendung bei gehörlosen (sowie resthörigen und schwerhörigen) Kindern und Jugendlichen keinesfalls geringer als in der Anwendungspraxis bei normalhörenden Pbn. Der *Standardmeßfehler* liegt nach unseren jüngsten Berechnungen an Hand von über 1000 PMT-Protokollen zwischen 1 und 2 Rohpunkten. Bezogen auf die *PMT-Normen für Hörgeschädigte* (vgl. Tab. Ia bis Ie) beträgt der Meßfehler in den Altersgruppen 5;0 bis 9;11 bzw. 15;0 bis 19 + sowie auf der Grundschulstufe (Klassen 1 bis 4) und in der Hauptschulklasse 9 jeweils rd. 1 RP, in den Altersgruppen 10;0 bis 14;11 sowie auf der Hauptschule (Klassen 5 bis 8) jeweils 2 RP. Für die PMT-Anwendung bei hörgeschädigten Real-, Wirtschafts- und Berufsschülern errechneten wir wiederum einen Meßfehler von 1 RP. Auf Standardwerte bezogen beträgt der Meßfehler des PMT durchweg ca. 2 IQ-Punkte (s.S. 66f.).

Die *Validität* des PMT ist in bezug auf die Erfassung der Allgemeinbegabung erstaunlich hoch. So korreliert die PMT-Leistung mit den wesentlich aufwendigeren Binet-Verfahren und dem WISC je nach Untersuchungssample zwischen 0.55 und 0.75, vereinzelt auch höher; bezüglich der Subtestvergleiche (im WISC bzw. HAWIK) ergaben sich sehr enge Zusammenhänge zwischen PMT und MT, wenngleich der PMT im allgemeinen enger mit den Binettests korreliert (vgl. Raven 1956, 1958; Wewetzer 1964). Nach Seifert (1960), der unter anderem die PMT-Ergebnisse mit diversen Lehrerurteilen verglich, besteht zwischen Lehrer- und Testaussage bei normalhörenden Schülern ein höherer Zusammenhang als bei gehörlosen Pbn, wo lediglich mittlere Koeffizienten um 0.5 ermittelt werden konnten. Wir selbst fanden bei zahlreichen PMT-Untersuchungen Gehörloser und Schwerhöriger Übereinstimmungskoeffizienten in der Relation PMT – SON zwischen 0.53 und 0.89 sowie von 0.53 bis 0.80 in der Relation PMT und HAWIK-Handlungsteil, was auf befriedigende bis gute, teilweise sogar sehr gute PMT-Validität hinweist (Heller 1967b, S. 227 f.). Im Gegensatz dazu fand Evans (1966) in seiner Vergleichsuntersuchung mit dem PMT und WISC-Performance an 100 taubstummen und resthörigen (englischen) Schülern im Alter von 6 bis 15 Jahren nur sehr mäßige Zusammenhänge; die Korrelationskoeffizienten lagen bei 0.27 und 0.44, wobei möglicherweise Stichprobenartefakte eine Rolle gespielt haben. Überhaupt sollte man bei der Beurteilung dieser oder anderer Validitätskoeffizienten (stärker als es vielfach geschieht) stets die Größe versus Kleinheit sowie die Repräsentanz versus Irrepräsentativität der (vieler) Untersuchungssamples berücksichtigen. Gerade im sonderpädagogischen und/oder klinischen Anwendungsbereich haften entsprechenden teststatistischen Aussagen aus begreiflichen Gründen (s.S. 76 oben) nicht selten Stichprobenmängel an. Auch Wewetzer (1964, S. 214) meint in bezug auf entsprechende PMT-Untersuchungen: „Vergleichsuntersuchungen zu anderen Intelligenztests zeigen verschiedene Ergebnisse . . . . Diese Maße hängen weitgehend von der Organisation der Stichprobe hinsichtlich Alter und Niveau ab.“ Auf der Grundlage spezifischer PMT-Normen für *Hörgeschädigte* (vgl. Tab. Ia bis Ie im Anhang) errechneten wir jüngst bei einer größeren Schülerzahl mittlere Validitätskoeffizienten von 0.7 in der Relation PMT – SON sowie 0.8 in der Relation PMT und HAWIK-Handlungsteil. Diese Ergebnisse dürften für bundesrepublikanische bzw. deutschsprachige Verhältnisse repräsentativ sein. Auch zwischen PMT- und Schulleistung (Gehörloser und

Schwerhöriger) war der ermittelte Zusammenhang mit  $r = 0.64$  recht beachtlich; dabei korrelierten PMT und Rechenzensur mit  $r = 0.68$  am höchsten, PMT und Kopfnoten (Mitarbeit und Fleiß mit  $r = 0.29$  erwartungsgemäß am niedrigsten (siehe unten Kap. III, 2b).

Die Original-Normen zum PMT beziehen sich auf ca. 7000 englische Pbn und sind als Prozentränge (PR) für 6 – 14jährige in Halbjahresschritten, für 20 bis 65jährige in 5 Jahresintervallen angegeben. Deutsche Eichwerte sind genauso wenig wie hörgeschädigtenspezifische PMT-Normen bekannt, obwohl der Raven-Test gerade hier relativ oft angewendet wird (unter Benutzung der englischen Normen für Normalhörende). Auf unsere Initiative hin erstellten deshalb 1966 Brunner, Lichtenberger und Schwinger auf der Basis von ca. 200 PMT-Protokollen hörgeschädigter Schüler erste – vorläufige – PMT-Normen für Hörgeschädigte (vgl. Heller 1967b, S. 231 ff.). Dank der Kooperationsbereitschaft vieler Lehrerkollegen an Gehörlosen- und Schwerhörigenschulen in der BRD konnte diese Informationsbasis inzwischen beträchtlich erweitert werden, so daß wir jetzt in der Lage sind, auf der Grundlage von über 1200 PMT-Protokollen für deutschsprachige Verhältnisse einigermaßen repräsentative PMT-Normen für Hörgeschädigte anzubieten. Diese finden sich als IQ-Normen für 5 – 19jährige bzw. 1. bis 12. Klassen in Tab. Ia bis 1e im Anhang dieses Buches abgedruckt.

**Testanwendung und Kritik:** Der PMT stellt ein ökonomisches und brauchbares Verfahren zur Erfassung der allgemeinen Intelligenz dar. Als nonverbaler Test ist er nicht nur für den sonderpädagogischen und klinischen Anwendungsbereich (z.B. für die Intelligenzdiagnose Hör- und/oder Sprachgeschädigter) interessant, er hat sich auch bereits im Rahmen der Begabungs- und Schuleignungsermittlung in der Regelschule, etwa als nonverbales Intelligenzdiagnostikum bei sprachlich oder überhaupt sozio-kulturell deprivierten Probanden, als nützlich und brauchbar erwiesen (z.B. Aurin 1966). Für Blinde liegt eine amerikanische PMT-Adaption vor (vgl. Rich 1963).

Für die Anwendung der PMT bei hirnorganisch bedingter Leistungsschwäche ist allerdings eine gewisse Vorsicht am Platze – wegen der allgemeinen Gestaltsschwäche von Hirnorganikern (vgl. Raven 1956, Wewetzer 1964 u.a.). Ebenfalls scheint der PMT nur bedingt für die Lernbehindertendiagnose (selection) geeignet zu sein. So differenziert er nach einer Untersuchung von Klauer (1964) nur mäßig zwischen Volks- und Hilfsschülern; gleichwohl hält derselbe Autor den PMT für heilpädagogische Zwecke als „besonders geeignet“ und schlägt hierzu eine deutsche Eichung vor. Weitere Untersuchungsergebnisse zum PMT berichten wir in Kap. III, 2b unten.

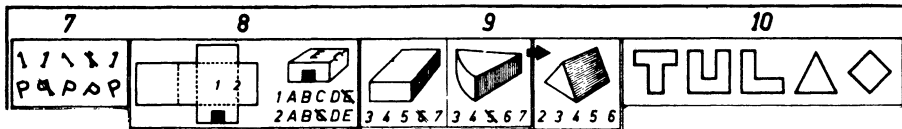
## 2. Differentielle Intelligenztests

Das **Leistungsprüfsystem (LPS)** von W. Horn (Verlag Dr. Hogrefe)

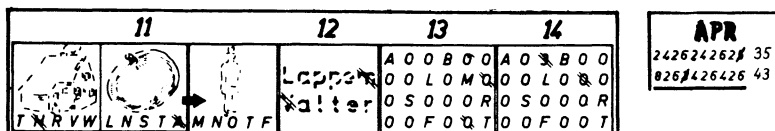
**Testaufbau:** Das LPS ist in Anlehnung an die Primary Mental Abilities (PMA) Testreihe von Thurstone 1962 von W. Horn entwickelt worden und ist „unter Einschränkungen jenes Testsystem in Deutschland, das dem Ziel eines differentiellen Fähigkeitstests am nächsten kommt“ (Pawlik 1970, S. 183). Die Testbatterie besteht aus 15 Subtests, zu denen hier jeweils die Übungsaufgaben abgebildet sind. Demnach werden in den LPS-Tests 1 und 2 Hauptwörter im Singular dargeboten, wobei jedes Wort (Item) *einen* Druckfehler enthält, der vom Pb durchzustreichen ist. Die Tests 3 und 4 repräsentieren Klassifikations- und Analogietests bzw. Aufgaben zum Reihenfortsetzen; das nicht passende Glied einer jeden (Figuren-versus Zahlen- oder Buchstaben-)Reihe ist wieder zu markieren. LPS-Test 5 ist ein Anagramm-

1	2	3	4	5	6
K r i d e	K f e i f e	○ ○ ○ ○ ● ○ ○ ○	2 2 2 2 2 2 2 2	G X W E R	F f F o o
T o l l e r	S c h l e	+   +   +	a b a b a b a a	C K E R X	K, k A u t

test; die richtige Reihenfolge der durcheinandergewürfelten Buchstaben ist zu erraten und der Anfangsbuchstabe des richtigen (herausgefundenen) Lösungswortes zu markieren. Im LPS-Subtest 6 sollen in bestimmten Zeiteinheiten möglichst viele Wörter zu vorgegebenen Anfangsbuchstaben gefunden werden. In Subtest 7 sind spiegelbildlich gedrehte Ziffern und Buchstaben aus plan verschobenen herauszufinden. Subtest 8 enthält Aufgaben zur Oberflächenabwicklung; der Pb muß vorgegebene Oberflächennetze in der Vorstellung falten und die Ziffern der Netze den Buchstaben der dreidimensional abgebildeten Körper zuordnen. Bei Subtest 9 muß die Flächenzahl vorgegebener Körper angegeben werden.



In Subtest 10 sind bestimmte, in ablenkende Gesamtfiguren eingebettete Figuren-Items zu identifizieren. Subtest 11 ist als Gestaltergänzungstest aufgebaut; der Anfangsbuchstaben der betr. Gegenstandsbezeichnung ist wieder durchzustreichen. Ähnlich sind im Subtest 12 unvollständige (jetzt Wort-)Gestalten zu ergänzen bzw. die Rechtschreibfehler zu markieren.



Die Subtests 13 und 14 verlangen die Durcharbeitung diverser Buchstaben- und Ziffernreihen, wobei bestimmte Regeln beachtet werden müssen. Subtest 15 repräsentiert eine Arbeitsprobe (APR); analog zum Pauli-Test sind hier Additionen von 10 einstelligen Zahlen in bestimmten Zeiteinheiten auszuführen.

Jeder der 14 Subtests im LPS enthält 40 Aufgaben, der gesamte Test umfaßt somit 560 Items (ohne APR). Die *Auswertung* des LPS, d.h. die Auszählung der Rohscores (Plus- und Fehlerpunkte), erfolgt mittels Lochschablonen. Die Rohpunktsummen werden getrennt für jeden Subtest in ein Diagramm (Plus-Spalte) eingetragen; abschließend wird ein Gesamt-Summenscore gebildet und in die GL-Spalte (Gesamtleistung) übertragen (vgl. Grafik 5 unten), was streng genommen mit den Konstruktionsprinzipien eines differentiellen Fähigkeitstests unvereinbar ist. Nach der Rohwert-Transformation in Centil-Werte ist eine *Profilzeichnung* möglich. Die gesamte LPS-Durchführung dauert zwei Unterrichtsstunden (90 Minuten); der Test kann als Gruppen- oder Individualverfahren administriert werden. Für die LPS-Auswertung benötigt eine eingearbeitete Kraft 10 bis 15 Minuten pro Pb-Protokolle (Form A 1 und A 2 versus Parallelform B 1 und B 2). Häufig werden von der Arbeitsprobe nur die ersten beiden – von insgesamt 10 – Aufgabenreihen durchgeführt.

**Erfassungsdimensionen:** W. Horn (Testmanual, S. 6) ordnet den einzelnen Subtests im LPS folgende Primärfähigkeiten sensu Thurstone zu (vgl. auch unsere früheren Ausführungen auf S. 17ff. sowie Heller 1970a, S. 100f.):

- Subtests 1 + 2 Verbal Factor; Allgemeinbegabung oder „Bildungsstandard“; Rechtschreibkenntnisse; Wort- und Begriffsschatz (vermutlich ist der Faktor W beteiligt)
- Subtests 3 + 4 Reasoning Factor; allgemeine, relativ sprach- und bildungsunabhängige Denkfähigkeit; logisches und schlußfolgerndes Denken (sehr komplexer Faktor)
- Subtests 5 + 6 Wordfluency; Worteinfall und Wortflüssigkeit (mehr assoziative, relativ inhalts- bzw. sinnunabhängige sprachliche Reproduktionsleistung); bei Subtest 5 ist der Faktor W beteiligt

Subtest 7 – 10	Space 1 und 2; Closure 2; Technische Begabung; Technisch-konstruktives Denken, mehr theoretisch-technische Findigkeit
Subtests 11 + 12	Closure, Closure 1; Ratefähigkeit; visuelle bzw. optische Gestaltauffassung; Flexibilität der Gestaltwahrnehmung sensu French
Subtests 13 + 14	Perceptual Speed; Accuracy; Wahrnehmungstempo; Fehlerbemerken bzw. Richtigkeit des Reagierens
Subtest 1 – 14	Gesamtleistung (GL); Begabungsniveau
Subtest 15 (APR)	Number Factor; Rechenfertigkeit (nicht: Rechenfähigkeit); Leistungsmotivation; Konzentration, Ausdauer, Belastbarkeit; Güte der Arbeitsleistung (Arbeitsqualität)
Variable –13	Fehler-„Leistung“; Arbeitsqualität; Sorgfaltsstreben

In der Anwendung des LPS bei 13jährigen Volksschülern konnte Tent (1969, S. 92 ff.) folgende *psychologische Dimensionen* aufgrund eigener Faktorenanalysen bestimmen:

1. *Sprachliche Leistungsfähigkeit* (mit den höchsten – ihrem Gewicht zufolge bereits geordneten – Ladungen auf den LPS-Subtests 1, 2, 12, 5 und 6);
2. *Schlußfolgerndes Denken* (mit den höchsten Ladungen auf den LPS-Subtests 3, 4, 13, 8 und 15);
3. *Räumliches Vorstellen* (mit den höchsten Ladungen auf den LPS-Subtests 8, 11, 10 und 9);
4. *Leistungsmotiviertheit* (mit den höchsten Faktorladungen in den LPS-Subtests 14 und 15);
5. *Wiedererkennen von Symbolen* (mit den höchsten Ladungsgewichten auf den LPS-Subtests 7, 12 und 5);
6. *Schulisches Lernen* (mit den höchsten Ladungsgewichten in den LPS-Subtests 15, 9 u. 6).

„Zusammenfassend läßt sich das LPS für die untersuchte Altersstufe als ein *psychologisch gut gegliederter und reichhaltiger Indikator des Leistungspotentials* kennzeichnen... Der Umstand, daß die von Horn postulierte Faktorenstruktur des LPS *innerhalb der Batterie* nur teilweise reproduziert werden konnte, ist demgegenüber unerheblich“ (loc. cit. S. 100 f.). Auf die Funktion des LPS als *Schuleignungsdiagnostikum* werden wir weiter unten (vgl. Kap. III, 1) zu sprechen kommen. In Parenthese sei hier noch ein Hinweis Tents erwähnt, wonach er für *Legastheniker* Ausfallserscheinungen in den LPS-Subtests 5, 7 und 12 erwartet.

**Testgütekriterien:** Die *Objektivität* der LPS-Durchführung und Testauswertung ist bei korrekter Einhaltung der Instruktions- und Auswertungsbestimmungen sehr gut.

Die *Reliabilität* der einzelnen Subtests liegt zwischen 0.90 und 0.99 (Halbierungszuverlässigkeit), die des Gesamttests bei 0.99 (korrigierte Splithalf-Reliabilität). In der LPS-Anwendung bei 10jährigen Pbn (4. Grundschulklasse) fand Tent (1969, S. 87 ff.) Halbierungskoeffizienten, die nur geringfügig unter den oben genannten, von Horn bei Erwachsenen gefundenen, Werten liegen, nämlich Koeffizienten zwischen 0.88 und 0.97, bezogen auf LPS-GL sogar 0.98, die andeuten, „daß *das LPS auch für die Kindergruppe eine beachtlich hohe innere Konsistenz seiner Meßfunktion behält*“ (loc. cit.). Die Retest-Reliabilitätskoeffizienten, die Tent bei einem Sample von 13jährigen Schülern nach 32monatigem Testintervall ermittelte, lagen im Mittel der 15 Subtests immerhin noch bei 0.6, bezogen auf LPS-GL bei 0.83; ebenso ergab sich eine genügend hohe Paralleltestreliabilität in den Untersuchungen Tents. Die Subtests korrelieren durchschnittlich um 0.45 miteinander, was durch teilweise identische Faktorenladungen in dieser Höhe bedingt sein mag. Tent fand allerdings niedrigere Interkorrelationen der LPS-Subtests im 5. und 6. Schuljahr. Mit der Gesamtleistung korrelieren die LPS-Subtests im Mittel um 0.7, was relativ hohe Untertestbeteiligung an LPS-GL bedeutet. Der *Meßfehler* des LPS beträgt nach unseren Berechnungen 0,3 C und bezieht sich auf die LPS-Anwendung im 4. Grundschuljahr sowie die LPS-Durchführung auf der Sekundar-

stufe I (Hauptschule, Realschule, Gymnasium); siehe auch die früheren Ausführungen auf Seite 68f.

Eine unerläßliche Voraussetzung für die *Profilinterpretation* eines differentiellen Intelligenztests ist seine *Profilreliabilität*, die gewöhnlich niedriger ausfällt als die Reliabilität des Gesamttests versus der Untertests. Die Profilreliabilität ( $\text{prof}_{\text{rft}}$ ) des LPS liegt nach unseren Berechnungen – je nach Untersuchungspopulation und/oder Reliabilitätskriterium (z.B. Halbierungs- oder Retest-Reliabilität) – zwischen 0.91 und 0.47; bei Grundschulern der 4. Klasse beträgt die Profilreliabilität des LPS 0.87, bei Sekundarschülern (Hauptschülern, Realschülern, Gymnasiasten) 0.91 bzw. 0.66 (bei nicht minderungskorrigierten Reliabilitätskoeffizienten in der  $\text{prof}_{\text{rft}}$ -Formel). Weitere Ergebnisse unserer teststatistischen Berechnungen zum LPS werden unten in Kap. III, 1 referiert.

Zur *Validität* des LPS liegen erst vereinzelt brauchbare Untersuchungsergebnisse vor. Die von Horn (1962, S. 24) mitgeteilten  $r_{\text{tc}}$ -Koeffizienten variieren zwischen 0.61 (in der Relation LPS 1 – 6 und Lehrerurteil über Intelligenz) und 0.17 (in der Relation LPS 7 – 10 und Lehrerschätzurteil); in bezug auf LPS-GL und Lehrerurteil wurden Validitätskoeffizienten um 0.5 ermittelt. Diese Ergebnisse beziehen sich auf Stichprobenerhebungen Horns im 4. Grundschuljahr (N = 175) sowie Untersuchungssamples auf der Volksschuloberstufe (Hauptschule) und Realschule (N = 354). Je nach Untersuchungszweck und Pbn-Gruppe liegen also die Validitätskoeffizienten des LPS – wie bei allen vergleichbaren Intelligenztests – höher oder niedriger, was aber nicht unbedingt dem LPS-Diagnostikum angelastet werden muß, sondern auch an der Wahl des Außenkriteriums liegen kann. Beispielsweise läßt sich der niedrige Zusammenhang von LPS 7 – 10 und Lehrerurteil dahingehend interpretieren, daß in den betr. LPS-Subtests andere (Intelligenz-) Faktoren erfaßt werden, als sie für die Schulleistung relevant resp. für die allgemeine Begabung charakteristisch erscheinen. Tent (1969, S. 110 ff.), der an Hand seines Untersuchungsmaterials die LPS-Gesamtleistung zur Schulleistung in der 5. Klassenstufe des Gymnasiums, der Real- und Hauptschule in Beziehung setzte, errechnete entsprechende Validitätskoeffizienten zwischen 0.5 und 0.6, wobei allerdings LPS-GL mit den Einzelfachzensuren nur zwischen 0.31 (Naturkunde bzw. Geschichte) und 0.55 (Raumlehre) korreliert war (Mediane biserieller Korrelationen). Die *Prognosegültigkeit* des LPS hinsichtlich der Schulbewährung auf den weiterführenden Bildungseinrichtungen kann nach Tent als zufriedenstellend bezeichnet werden. „Faßt man die Einteilung in Volksschulen, Realschule und Gymnasium als Trichotomisierung einer kontinuierlichen, hinreichend normal verteilten Variablen auf, die man „quantitativ abgestufte schulische Leistungsforderung“ nennen könnte, läßt sich der Zusammenhang zwischen LPS-Gesamtpunktwert und dem Schulbesuch im 5. Schuljahr durch einen triserialen Korrelationskoeffizienten (N. Jaspens 1946) von  $r_{\text{tris}} = 0.59$  wiedergeben“ (loc. cit.). Ähnliche – positive – Erfahrungen mit dem LPS wurden u.a. von Aurin (1966) Heller 1970 a, 1970 b), Allinger & Heller (1972) berichtet. Das LPS kann somit als ausreichend valide im Hinblick auf die Begabungsermittlung resp. Schuleignungsdifferenzierung angesehen werden.

Für das LPS liegen *Altersnormen* – in C-Werten ( $M_C = 5$ ;  $s_C = 2$ ) – ab 9 Jahren (für GL) bzw. ab 10 Jahren (für die Subtests) vor, die zunächst in Jahresabständen, vom 21. bis zum 50. Lebensjahr in 10-Jahresschritten tabelliert sind.

**Anwendungsbereiche:** Das LPS stellt ein objektives, zuverlässiges und gültiges Intelligenz-Leistungsdiagnostikum dar, das sowohl die allgemeine Begabung als auch differentielle (teilweise relativ bildungs- und sprachunabhängige) Fähigkeiten erfaßt. Es empfiehlt sich gleichermaßen für Intelligenzdiagnosen und Schuleignungsprognosen.

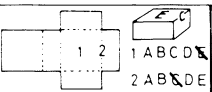



**Kritik:** Die Vorteile des LPS liegen einmal in seiner Ökonomie (der Testdurchführung und -auswertung) sowie in den erwiesenen Testgütekriterien, zum andern repräsentiert das LPS ein brauchbares Konzept zur Testung wichtiger Primärfähigkeiten. Wie kein anderes bisher behandeltes Verfahren eignet sich der Horn-Test zur *Profilanalyse*. Freilich liegen hierin auch noch gewisse Schwächen, die sich etwa in den vergleichsweise (im Hinblick auf die Konstruktionsprinzipien eines Profilverfahrens immer noch zu) hohen Testinterkorrelationen

nen sowie der beachtlichen Korrelation der Subtests mit der Gesamtleistung ausdrücken. Ferner müßten die Subtests faktorentreiner gemacht sowie spezielle Validierungsstudien im Sinne einer optimaleren Profilauswertung in Angriff genommen werden. Trotzdem hat sich das LPS als faktorisierte Begabungstest bei der diskriminanzanalytischen Bearbeitung von Klassifikationsproblemen im Bildungswesen recht gut bewährt (vgl. Heller 1970a, Allinger & Heller 1972).

**Das Prüfsystem für Schul- und Bildungsberatung (PSB) von W. Horn (Verlag Dr. Hogrefe)**

Das PSB ist aus dem LPS entstanden und stellt praktisch eine Kurzform desselben dar. Aufgrund der Befunde umfangreicher Diskriminanzanalysen beim LPS im Rahmen der baden-württembergischen Bildungsberatung (vgl. Aurin 1966, S. 147 ff.) wurden hier diejenigen LPS-Subtests zum sog. PSB zusammengefaßt bzw. ausgewählt, denen im Hinblick auf die Begabungs- und Schuleignungsdifferenzierung die beste diakritische Funktion zukommt.

**Testaufbau:** Die PSB-Subtests 1 + 2 sowie 3 und 4 sind praktisch, d.h. von kleineren Veränderungen abgesehen, mit den LPS-Subtests 1 + 2 sowie 3 und 4 identisch und brauchen hier nicht mehr beschrieben zu werden. PSB-Subtest 5 entspricht LPS-Subtest 6. Die

5	6	7	8	9	10
LI-Lernen	LSP-3				
PP-Bach	Kalter	1 2	1 ABCD 2 ABDE	73468293	7348293
				64372859	6437859

PSB-Subtests 7 und 8 sind mit den LPS-Subtests 8 bzw. 10 identisch. Lediglich die Subtests 9 und 10 im PSB wurden gegenüber der Arbeitsprobe bzw. dem Subtest 14 im LPS stärker verändert. Hingegen ist der PSB-Subtest 6 wiederum mit dem LPS-Subtest 12 identisch. Siehe die abgebildeten Übungsbeispiele zu den PSB-Subtests 5 bis 10; bezüglich der Übungsbeispiele zu den PSB-Subtests 1 bis 4 vgl. oben S. 135.

**Erfassungsdimensionen:** Analog zu den entsprechenden LPS-Subtestdimensionen formuliert Horn (1969, S. 18) für den PSB folgende Erfassungsdimensionen oder Primärfähigkeiten sensu Thurstone (in Klammern die PSB-Subtests, auf denen die betr. Faktoren laden):

Verbalität	(1+2, 6, 5)
Denkfähigkeit	(4, 3)
Technische Begabung	(7, 8)
Closure	(8)
Wahrnehmungstempo	(10)

**Testgütekriterien:** Im großen und ganzen gelten hier dieselben Gütekriterien wie beim LPS (siehe oben). Das arithmetische Mittel aller Subtestinterkorrelationen im PSB beträgt nach unseren Berechnungen, bezogen auf die von Horn adaptierten LPS-Werte für den PSB (Horn 1969, S. 18), 0,51 beziehungsweise, auf die von Aurin bei 907 Viertklässkindern ermittelten PSB-Interkorrelationen bezogen, 0,41, was die Homologie der Strukturen beider Horn-Tests unterstreicht. Die durchschnittliche Halbierungszuverlässigkeit aller Subtests im PSB liegt bei 0,92 (Viertklässkindern) und 0,96 (Erwachsene). Dementsprechend ermittelten wir eine PSB-**Profilreliabilität** für die Anwendung im 4. Schuljahr von  $\text{prof}_{\text{TT}} = 0,86$  (nach den Hornschen Interkorrelationen) und  $\text{prof}_{\text{TT}} = 0,84$  (nach den Aurinschen Interkorrelationen). Weitere Ergebnisse teststatistischer Berechnungen zum PSB finden sich in Kap. III, 1 u. 2 unten. Zum PSB liegen wiederum **Altersnormen** für 9. bis 18. bzw. 20. Lebensjahr vor, für die Altersgruppen 9 bis 11 jeweils in Vierteljahresintervalle unterteilt. Die Centil-Normierung des LPS wurde auch hier beibehalten, d.h. die Normen sind in C-Werten angegeben.

**Anwendungsbereiche:** Das PSB empfiehlt sich besonders in der Grundschule (Klasse 4) sowie auf der Sekundarstufe I überall dort, wo intelligenzdiagnostische und begabungs- oder schuleignungsprognostische Fragen vorliegen. Seine Durchführung beansprucht die Hälfte der LPS-Bearbeitungszeit (1 Stunde). Die PSB-Zuverlässigkeit liegt zumindest auf der Grund- und

Hauptschule nur geringfügig unter der LPS-Reliabilität. Bei Gymnasiasten und Realschülern der Oberstufe wurde im Rahmen von Begabungsuntersuchungen mitunter eine gegenüber dem LPS verminderte Reliabilität festgestellt, weshalb dort dem LPS-Einsatz Vorrang gebührt.

**Kritik:** Die bei der Erörterung des LPS vorgetragenen Kautelen versus Verbesserungsvorschläge gelten sinngemäß auch hier.

### 3. Tests und Verfahren zur Erfassung spezieller Fähigkeits- und Verhaltensmerkmale

#### a) Individualtests bzw. Einzelverfahren

Der **Benton-Test** von A.L. Benton in der deutschen Bearbeitung von O. Spreen (Huber Verlag)

Die erste amerikanische Fassung des Benton-Testes erschien 1945 unter der Bezeichnung „The Visual Retention Test“ (vgl. Benton 1945). Inzwischen liegt die dritte, erweiterte Auflage des von Spreen erstmals 1958 ins Deutsche übertragenen „Revised Visual Retention Tests“ vor (Benton 1961, 1968), die auch unserer Besprechung zugrunde liegt.

**Testaufbau:** Der Benton-Test besteht aus zwei Formen, der sog. *Zeichenform* (ursprüngliche Testvorlage) und der sog. *Wahlform* (multiple Choice type). Die *Zeichenform* setzt sich aus den Paralleltestserien C, D und E mit je 10 Testitems, die in den Vorlagenheften 1 und 2 enthalten sind (vgl. Benton 1961), zusammen. Die *Wahlform* besteht aus zwei Paralleltestserien (Form F im Aufgabenheft 1; Form G im Aufgabenheft 2) zu je 15 Testitems. Während die *Zeichenform* (davon meist nur 1 Aufgabenserie zu 10 Items) als Standardtest fast immer verwendet wird, ist die *Wahlform* mehr für motorisch gestörte Pbn (z.B. Kinder mit zerebralen oder Halbseitenlähmungen) gedacht. Oft werden auch beide Formen kombiniert – zur besseren Kontrolle der Testergebnisse – eingesetzt. Die einzelnen Testformen und Aufgabenserien können unter variierten Instruktionsbedingungen appliziert werden, wobei Instruktion A (jede Itemkarte wird 10 Sekunden lang dargeboten) die Standard-Versuchsanordnung der *Zeichenform* und Instruktion M (ebenfalls 10 Sekunden Darbietungszeit) die Standard-Versuchsanordnung der *Wahlform* darstellt. Unsere im Appendix als Tab. IIa und IIb beigefügten *Hörschädigtennormen zum Benton-Test* basieren ausschließlich auf den genannten Standard-Versuchsanordnungen A und M. Die Benton-Testaufgaben ähneln sehr stark dem Bender-Gestalt-Test (vgl. Cohen 1964), doch müssen im Visual-Retention-Test die Figuren nicht direkt abgezeichnet, sondern aus dem Gedächtnis reproduziert werden. Beim Multiple-Choice-Verfahren müssen die Stimuluskarten aus jeweils 4 Wahlvorschlägen wiedererkannt, d.h. aus dem Gedächtnis zugeordnet werden. Die *Zeichenform* kann quantitativ *und* qualitativ (Fehleranalyse), die *Wahlform* nur quantitativ ausgewertet werden.

**Erfassungsdimensionen:** Cohen rechnet den Benton-Test den Zeichentests zur Prüfung der (allgemeinen?) *Intelligenz* zu, und als solcher wird er hier zuallererst erwähnt. Darüber hinaus ist der Benton-Test nach der Meinung des Testautors und einer Reihe anderer Experten zur *Erfassung hirnorganischer Störungen* geeignet, obschon die Ansichten dazu insgesamt ziemlich kontrovers sind (vgl. aber Benton 1968, S. 60 ff. und S. 93 ff.).

**Testgütekriterien:** Während die Objektivität der Testdurchführung gewährleistet erscheint, ist die *Objektivität der Testauswertung* keinesfalls im gewünschten Maße gesichert. Die Angaben zur Auswerterinterkorrelation streuen von 0.98 bis 0.66, was zum Teil gute und zum Teil unzureichende Auswertungsreliabilität bedeutet.

Die *Retest-Reliabilität* des Benton-Tests wird mit  $r = 0.85$  (*Zeichenform*) bzw. mit  $r = 0.80$  (*Wahlform*) angegeben und signalisiert somit ausreichende Testzuverlässigkeit im Sinne der Zeitstabilität.

Die *Validität* des Benton-Tests gibt Benton (1968, S. 53) mit  $r_{TC} = 0.70$  an. Nach einer Vergleichsuntersuchung von Batt (1967) bei tauben und hörrestigen Schülern sind die BT-Er-

gebnisse mit dem PMT von Raven durchschnittlich mit 0.35 korreliert, und zwar im Alter von 7;6 bis 10;5 mit  $r = 0.40$ , von 10;6 bis 13;5 mit  $r = 0.16$  und von 13;6 bis 17;0 mit  $r = 0.54$ . Mit dem Mosaiktest im HAWIK korreliert der Benton-Test bei den gleichen Stichproben mit  $CC_{\text{corr}} = 0.48$ , in der Altersgruppe der 7 bis 10jährigen Hörgeschädigten sogar mit  $CC_{\text{corr}} = 0.55$ ; Validitätskoeffizienten um 0.4 sind gerade noch akzeptabel, Koeffizienten um 0.6 deuten bereits recht gute Validität an. Die von Batt ermittelten Werte liegen aber deutlich unter den von Benton angegebenen, die sich freilich auf Normalhörende beziehen. Möglicherweise ist in den Untersuchungssamples von Batt eine höhere Anteilsquote hirnanorganisch gestörter resp. verdächtiger Pbn hier als Interferenzfaktor zu interpretieren.

*Normwerte* werden für Kinder vom 7. bis 13. Lebensjahr sowie für Erwachsene (14. bis 55. Lbj.) im Testmanual angegeben, wobei jeweils bestimmte Fehler-Punktzahlen den von Wechsler definierten IQ-Klassen zugeordnet sind. Ferner werden grobe Leitwerte klinisch auffälliger Gruppen benannt. Neben bestimmten Fehlerhäufigkeiten werden Drehungen, Perseverationen, Größenveränderungen sowie Fehlplacierungen als valideste Testmerkmale zerebralpathologischer Erscheinungen gewertet. „Es besteht auch eine signifikante Abhängigkeit zwischen dem BT und dem Lebensalter. Die Leistungshöhe des Tests steigt progressiv vom Alter von 8 Jahren an, bis eine gleichbleibende Höhe im Alter von 14 bis 15 Jahren erreicht wird. Diese Höhe wird von der ausgehenden Pubertät bis ins vierte Lebensjahrzehnt hin gehalten. Zwischen 40 und 50 Jahren tritt eine Abnahme der Leistungshöhe ein, die während der folgenden Lebensjahrzehnte weiter fortschreitet. So zeigen unsere Standardwerte ein Abfallen des durchschnittlichen Wertes für richtige Reproduktion um etwa einen Punkt für die Altersgruppe von 45 bis 54 Jahren und um zwei Punkte für die Altersgruppe von 55 bis 64 Jahren im Vergleich zu den jüngeren Altersgruppen“ (Benton 1968, S. 53). Siehe auch die im Anhang unseres Buches beigegebenen BT-Normen für Hörgeschädigte (Tab. II).

*Anwendungsbereiche:* Der Benton-Test kann als (nonverbaler) Intelligenztest und/oder zur Erkennung hirnanorganischer Störungen bzw. erworbener Intelligenzminderungen (bei einer Fehlerpunkte-Differenz zum jeweiligen Altersnormwert von mehr als 3 Punkten) in der sonderpädagogischen und klinischen Diagnostik, gelegentlich auch innerhalb der Regelschule, eingesetzt werden.

*Kritik:* Je nach Zusammensetzung der Untersuchungspopulation müssen dem Benton-Test gute bis mäßige, teilweise auch unzureichende Testgütekriterien bescheinigt werden. Unter entsprechenden Kautelen ist eine begrenzte Anwendung des Benton-Tests in der Schule bzw. Sonderpädagogik durchaus sinnvoll oder sogar erforderlich, etwa bei Verdachtsmomenten zerebralpathologischer Art. Als Intelligenzdiagnostikum steht der Benton-Test allerdings in seiner Effizienz den oben beschriebenen Intelligenztests um vieles nach. Für die Hör- und/oder Sprachgeschädigtendiagnostik ist er als nonverbaler Test interessant.

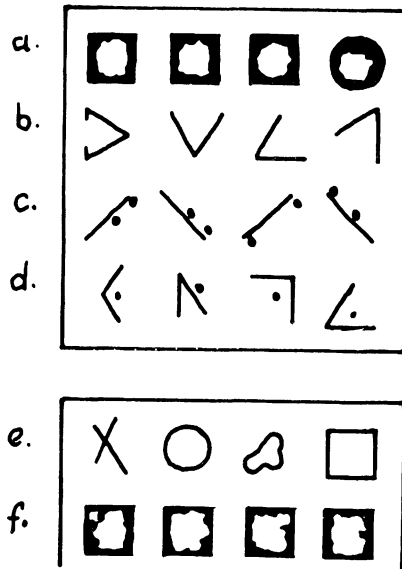
### Vocational Intelligence Scale for Adult Blind (VISAB) von W. R. Jones (TZ)

Die VISAB-Serie wurde im Rahmen eines Forschungsprojektes zur beruflichen Rehabilitation blinder Erwachsener an der Purdue Universität Michigan (USA) von W.R. Jones (1960) als Dissertation erstellt. Hierbei handelt es sich um einen nonverbalen, taktilen Intelligenztest, der in einer Kontrolluntersuchung von M. Strauß (1967) auf seine Verwendbarkeit hin bei (47 deutschen) blinden *Schülern* im Alter von 10;0 bis 15;11 Jahren erprobt wurde.

*Testaufbau:* Der VISAB besteht aus 6 Übungsaufgaben (siehe untenstehende Abbildung) und 43 Testitems mit unterschiedlichen geometrischen Figuren. Der Pb soll jeweils dasjenige Muster herausfinden, das die geringste Verwandtschaft zu den anderen Mustern einer Reihe aufweist. Die Figuren sind also nach dem Grad ihrer Verwandtschaft zu klassifizieren. Die Übungsaufgaben sowie die eigentlichen Aufgabenmuster wurden für die deutsche Untersuchung aus Plastik (im Tiefziehverfahren) hergestellt. Diese können den interessierten (Blinden-)Schulen gegen einen Unkostenbeitrag vom Institut für Sehgeschädigtenpädagogik der Pädagogischen Hochschule Heidelberg zur Verfügung gestellt werden. Detaillierte Hinweise



zur Instruktion des VISAB bei jugendlichen und erwachsenen Blinden sowie die notwendigen Auswertungskriterien (Schlüssel) finden sich in den Originalarbeiten von Jones bzw. – übersetzt – bei Strauß (1967, S. 34 ff.).



**Erfassungsdimensionen:** Der VISAB mißt die (allgemeine) *Intelligenz* (Denkfähigkeit) und ist nach Jones ein zuverlässiger Indikator zur Bestimmung des *Berufserfolges* Blinden.

**Testgütekriterien:** Die *Objektivität* des VISAB kann als gut bezeichnet werden. Die *Reliabilität* liegt nach Kuder & Richardson (Konsistenzanalyse) bei  $r = 0.91$ . Die *Validität*, gemessen als Übereinstimmung von VISAB und WAIS-Verbal, beträgt  $r_{tc} = 0.63$ ; außerdem wurden vier Kriterien beruflichen Erfolges Blinder aufgestellt: Berufsgruppe, Verdienst, Berufsrangordnung und Vorgesetzten-Urteil. Die entsprechenden Validitätskoeffizienten variieren zwischen 0.40 und 0.69 und unterstreichen somit nach Meinung des Testautors die berufliche Vorhersagegültigkeit bei Blinden. Strauß fand bei ihrer VISAB-Untersuchung blinder Kinder und Jugendlicher folgende  $r_{tc}$ -Werte zum HAWIK-Verbalteil:

VISAB – AW :	$r = 0.63$	VISAB – GF :	$r = 0.59$
VISAB – AV :	$r = 0.68$	VISAB – WT :	$r = 0.60$
VISAB – RD :	$r = 0.74$	VISAB – ZN :	$r = 0.44$

Die altersunabhängig ermittelten Koeffizienten liegen jeweils nur unbedeutend unter den oben angeführten, im Durchschnitt bei 0.61 (Strauß 1967, S. 29 f.). Mit den Schulleistungen (Zensuren) in Rechnen korreliert der VISAB mit  $r = 0.63$ , im Lesen mit  $r = 0.37$ , im Rechtschreiben mit  $r = 0.43$  und mit Formen/Bildhaftes Gestalten mit  $r = 0.61$ .

Die *Eichung* des VISAB steht praktisch noch aus. An Hand des deutschen Blindensamples berechnete Strauß (loc. cit., S. 27 f.) für drei Altersgruppen die *Mutungsintervalle* (Vertrauensbereiche auf dem 1%-Sicherungsniveau) für den jeweiligen Populations-Mittelwert (M):

Gruppe I (10;0 bis 11;11) : M (geschätzter Parameter) im Bereich von 15 RP und 24 RP

Gruppe II (12;0 bis 13;11) : M (geschätzter Parameter) im Bereich von 21 RP und 32 RP

Gruppe III (14;0 bis 15;11) : M (geschätzter Parameter) im Bereich von 15 RP und 33 RP

Genauere Normierungsdaten fehlen; die Standardisierung des VISAB steht somit in einem wichtigen Punkt noch aus. Für eine vorläufige Anwendung des VISAB empfehlen wir deshalb die Bestimmung von Rangplätzen und deren anschließende Transformation in T-Werte mit Hilfe der im Appendix stehenden Tabelle IX. Als Ausgangsbasis für die Gewinnung der Rangplätze sollten die betr. Klassengruppe, der der Pb angehört, oder – noch besser – mehrere Klassen des betr. Schuljahrs und Schultyps dienen.

**Anwendungsbereiche:** Der VISAB eignet sich unter bestimmten Kautelen (z.B. hinsichtlich der Eichung) zur nonverbalen Intelligenzbestimmung *blinder* Kinder und Jugendlicher resp. Erwachsener, wobei mehr Faktoren der Handlungsintelligenz, aber auch wichtige Denkfähigkeiten der allgemeinen Intelligenz, erfaßt werden. Inwieweit darüberhinaus berufsrelevante Eignungsmerkmale ermittelt werden können, bedarf meines Erachtens noch klärender – im deutschsprachigen Raum bislang fehlender – Untersuchungen. Ferner empfiehlt Jones, den VISAB in Ergänzung zum WISC- bzw. HAWIK-Verbalteil, quasi als Ersatz des HAWIK- bzw. HAWIE-Handlungsteils, durchzuführen. Hier wäre der VISAB besonders für die jüngeren Sekundarstufenschüler interessant und könnte möglicherweise die Funktion des HISAB (s.S. 123ff.) in der Verlängerung nach unten (10. bis 15. Lbj.) übernehmen.

**Testkritik:** Das VISAB-Diagnostikum ist – ähnlich wie der oben behandelte CFT oder PMT – ein *nonverbaler* Intelligenztest. In der vorliegenden Form als *taktiler* Handlungstest läßt er sich gut bei *blinden oder hochgradig sehbehinderten* Kindern ab dem 10. Lebensjahr sowie bei Jugendlichen und Erwachsenen durchführen. Allerdings stehen entsprechende Normierungsarbeiten noch aus, weshalb wir eine vorläufige Rangplatzbestimmung (mit evtl. T-Transformation) vorschlugen. Erhebungen zur Normengewinnung auf repräsentativer Basis sowie weitere Reliabilitätskontrollen (Retestungen) erscheinen im Hinblick auf eine verläßliche Intelligenzdiagnose mit Hilfe des VISAB vordringlich, zumal kein Test dieser Art für die Blindendiagnostik hierzulande zur Verfügung steht.

**Der Konzentrations-Verlaufs-Test (KVT) von D. Abels (Verlag Dr. Hogrefe)**

**Testaufbau:** Der KVT besteht aus einem Paket von Karten mit je 36 aufgedruckten zweistelligen Zahlen. Diese müssen möglichst schnell nach bestimmten Regeln sortiert werden. Bartenwerfer (1964, S. 392 f.) hebt folgende Eigenschaften des KVT hervor: „1. Die Reihenfolge der Darbietung der (60) Karten ist standardisiert. 2. Aus der besonderen Art der Fixierung des Sortierergebnisses durch den Pb ergibt sich in Verbindung mit der Konstanz der Kartenreihenfolge die zusätzliche Möglichkeit, den Verlauf der Leistungsqualität zu verfolgen. 3. Durch die Instruktion, die Testanlage und eine besondere Auswertungsart, die mit Hilfe einer Auflegeschablone vollzogen wird, lassen sich drei verschiedene Arten von Fehlern unterscheiden: das Übersehen einer Zahl pro Karte (einfache Fehler), das Übersehen von 2 Zahlen pro Karte (Doppelfehler) und das Hineinsehen einer nicht vorhandenen Zahl in eine Karte (Illusionsfehler). 4. Die Instruktion lenkt den Pb besonders in Richtung auf sorgfältiges Arbeiten („Mit dieser Aufgabe wird Ihre Sorgfalt bei der Durchführung einer Arbeit untersucht.“). 5. In den Normentabellen läßt sich neben einer getrennten Bewertung von Testzeit und Fehlern auch ein kombinierter Standardwert aus Arbeitszeit und -qualität ermitteln.“

**Erfassungsdimensionen:** Der KVT gehört zu den Tests, die „anhaltende Konzentration bei geistiger Tempoarbeit“ (sensu Bartenwerfer bzw. Düker), also ein wichtiges Merkmal der Sekundärintelligenz bzw. der Hilfs- oder Stützfunktionen der Intelligenz (sensu Mierke) erfassen sollen. Im allgemeinen korrelieren Konzentrationstests dieser Art (siehe unten noch KLT, Bourdon- und Pauli-Test) schwach (zwischen 0.0 und 0.4) bis stärker (zwischen 0.3 und 0.7) mit den üblichen Intelligenztests. Eingehend orientiert Bartenwerfer (vgl. S. 398 ff.) über die von ihm veranlaßten faktorenanalytischen Untersuchungen zur Gültigkeitsbestimmung des KVT.

**Testgütekriterien:** Der KVT verfügt über ausreichende *Objektivität* sowohl hinsichtlich der Testdurchführung als auch der Testauswertung. Die Untersuchungen zur *Reliabilität* lassen

indessen noch manche Fragen offen. Die Halbierungs-Zuverlässigkeit des KVT wird mit 0.67 angegeben (Abels 1961), was als zu niedrig angesehen werden muß. Bartenwerfer (loc. cit.) ermittelte Paralleltestreliabilitäten zwischen 0.59 und 0.40; lediglich Seyfried fand günstigere Reliabilitätskoeffizienten (zit. nach Bartenwerfer).

Zur *Validität* des KVT liegen neben den faktorenanalytischen Befunden Bartenwerfers (siehe oben) „(Vierfelder-)Korrelationen zwischen den KVT-Fehlern und dem Lehrerurteil über „Konzentrationsfähigkeit“ in der Höhe zwischen 0.60 und 0.83 bei Stichprobenumfängen zwischen 196 und 401“ nach Seyfrieds Untersuchungen an österreichischen Pbn vor (zit. nach Bartenwerfer 1964, S. 393 f.). Offiziell liegen bislang nur Erwachsenen-*Normen* zum KVT vor. Jedoch fand Seyfried in Österreich, „daß der Test hoch gesichert zwischen verschiedenen Altersklassen bei Kindern differenziert und auch zwischen gleichaltrigen Gymnasiasten, Hauptschülern, Nicht-Hochschulreifen und Hilfsschülern“ (loc. cit.). Martin, der 1961 umfangreiche KVT-Erhebungen an hörgeschädigten Kindern und Jugendlichen durchführte, berechnete entsprechende Normen für gehörlose und resthörige deutsche Schüler. Diese finden sich von uns zusammengestellt im Anhang dieses Buches (vgl. Tab. VIa und VIb) ebenso wie die von Seyfried an österreichischen Schülern ermittelten KVT-Normen (vgl. Tab. VIc und VIId).

*Anwendungsbereiche:* Der KVT eignet sich zur Erfassung der Konzentrationsfähigkeit etwa ab dem 10. Lebensjahr. Er ist sowohl in der Regelschule als auch in der Sonderschule, z.B. bei Hör-/Sprachgeschädigten, anwendbar (vgl. Tab. VI im Anhang). An Hand der quantitativen Auswertung (Zeitleistung versus Arbeitsorgfalt) und der qualitativen Fehler- bzw. Verlaufsanalyse können wichtige Komponenten der *Konzentrationsfähigkeit* sowie des *Arbeitsverlaufs* bestimmt werden.

*Testkritik:* Allgemein wünschte man sich eine bessere Zuverlässigkeit der KVT-Meßergebnisse, insbesondere hinsichtlich der qualitativen Beurteilung (bei Kindern *und* Erwachsenen). Die mangelnde (Fehler-)Reliabilität scheint freilich ein Charakteristikum sämtlicher Konzentrationstests (vgl. noch KLT und Pauli-Test) zu sein.

### **Der P-I-Test. Ein kombinierter Persönlichkeits- und Interessentest (PIT)**

von E. Mittenecker & W. Toman (Verlag A. Sengl, Wien)

„Der PIT ist der erste mehrdimensionale Persönlichkeitsfragebogen in deutscher Sprache. Er ist mit einem Interessentest kombiniert, von dem erwartet wird, daß er die persönlichkeitsdiagnostische Verwertung des Gesamtprofils erhöht“ (Pawlik 1970, S. 239). Hier interessiert jedoch der PIT vor allem wegen des Interessenteils, der bereits ab dem 16. Lbj. (sonst erst ab dem 18. Lbj.) anwendbar ist und von uns im Rahmen von Begabungsuntersuchungen auf der Sekundarstufe (Klassen 10 und 12) erprobt wurde.

*Testaufbau:* Der PIT besteht aus 120 Feststellungen (Persönlichkeitsteil) und 94 ipsativen Items (Interessenteil), die jeweils auf Karten gedruckt sind. Alle Feststellungen bzw. Testitems können wahlweise mit „STIMMT“, „STIMMT NICHT“ oder „WEDER NOCH“ („WEISS NICHT“) beantwortet werden. Die 9 Persönlichkeitskalen (A bis I) sowie die Interessenskala 0 sind bipolar angeordnet, die übrigen 8 Interessenskalen unipolar (vgl. Mittenecker 1964):

- A: Selbstkritik versus Mangelnde Selbstkritik (+A vs. -A)
- B: Soziale Einstellung versus Unsoziale Einstellung (+B vs. -B)
- C: Extraversion versus Introversion (+C vs. -C)
- D: Nicht neurotisch versus Neurotisch (allgemeine Fehlanpassungskategorie) (+D vs. -D)
- E: Nicht manisch versus Manisch (+E vs. -E)
- F: Nicht depressiv versus Depressiv (+F vs. -F)
- G: Nicht schizoid versus Schizoid (+G vs. -G)
- H: Nicht paranoid versus Paranoid (+H vs. -H)
- I: Vegetativ stabil versus Vegetativ labil (+I vs. -I)

O: Landleben versus Stadtleben (+O vs. -O)	} Interessenteil
P: Handwerklich	
Q: Wissenschaftlich	
R: Verrechnung, Verwaltung	
S: Umgang mit Menschen in Geschäft und Wirtschaft	
T: Bildende Kunst	
U: Sprachlich, literarisch	
V: Musikalisch	}
W: Soziale Berufe	

Bei den weiter unten referierten Untersuchungen zum PIT wurden die Skalen A, C und O bis W zu einem 8seitigen (Interessen-)Fragebogen mit insgesamt 124 Items zusammengefaßt. In dieser Form kann der PIT auch als Gruppenverfahren appliziert werden.

**Erfassungsdimensionen:** Die Skala A dient einem doppelten Zweck, nämlich als Maß für die Zuverlässigkeit der Selbstbeurteilung sowie als Kriterium für die Zuverlässigkeit des Fragebogenresultates überhaupt (Lügenscore). Die – intendierten – Erfassungsdimensionen der übrigen Skalen gehen aus deren Bezeichnungen hervor. Inwieweit die *Persönlichkeitsskalen* diesem Anspruch genügen, ist nicht so leicht abzuschätzen. Immerhin konnte Saltzman (zit. nach Pawlik 1970, S. 241) nachweisen, daß die I-Skala zwischen einer Gruppe von hospitalisierten vegetativ gestörten Pbn und einer Gruppe von Sportlern bei Interkorrelationen zwischen 0.0 und 0.7 in der Mehrzahl der Fälle brauchbar differenzierte. Aufgrund faktorenanalytischer Untersuchungen konnte Janouschek (zit. bei Pawlik) 3 bis 4 interpretierbare P-Faktoren, d.h. genausoviele Faktoren wie der wesentlich umfangreichere MMPI (vgl. Literaturverzeichnis), ermitteln. Und Neugebauer schließlich schlug einen „Abweichungsindex vor, der die Summe aller negativen Werte in den Skalen D bis I plus die Hälfte der „Weiß nicht“-Antworten ist. Dieser Index soll mit einem Korrekturfaktor gewogen werden, der aus A berechnet werden soll: je geringer die Selbstkritik, desto größer die Gewichtung. Dieser Index differenziert zwischen Normalen und klinischen Gruppen, nicht aber zwischen den klinischen Gruppen“ (Pawlik, loc. cit.). Bezüglich des *Interessenteils* ist hier ein Hinweis Mitteneckers von Bedeutung. „In einer unveröffentlichten Arbeit verglich Mayer (1957) die Durchschnittsinteressenprofile von 15 Gruppen von männlichen (N = 349) und 12 Gruppen von weiblichen (N = 218) Maturanten mit verschiedenen „Idealberufswünschen“. Die Ergebnisse bestätigten ausnahmslos die inhaltliche Validität des Interessentests“ (loc. cit.). Wir selbst fanden bei einer deutschen Primanerstichprobe (N = 212) kanonische Korrelationskoeffizienten von 0.40 bis 0.01 zwischen PIT (Interessenteil) und BIT, was darauf hindeutet, daß im PIT teilweise andere Interessen erfaßt werden als im BIT von Irle (vgl. Heller 1970a, S. 103 f.).

**Testgütekriterien:** Die *Objektivität* des PIT ist gewährleistet, wenngleich subjektive Verfälschungen seitens des Pb trotz der „Lügenskala“ nicht ganz auszuschließen sind. Die *Retest-Reliabilität* der Persönlichkeitsskalen liegt bei 0.87 (nach 24 Stunden) bzw. 0.76 (nach 3 Monaten); mit Ausnahme der E-Skala, wo nur Stabilitätskoeffizienten von 0.6 bzw. 0.5 gefunden wurden, ist der PIT in seinem Persönlichkeitsteil somit ausreichend zuverlässig. Für den Interessenteil gelten nach Mittenecker Retest-Zuverlässigkeitskoeffizienten um 0.89 (nach 24 Stunden) bzw. 0.82 (nach 3 Monaten). Die in der Literatur berichteten Untersuchungsergebnisse zur *Validität* des PIT schwanken je nach Skala und/oder Anwendungspopulation zwischen 0.64 und 0.93 (PIT – MMQ u.ä.) resp. zwischen 0.54 und 0.71 (PIT – Rorschach-Test). Zur *inhaltlichen* und *kriterienbezogenen* Gültigkeit des Interessenteils (0.4) äußerten wir uns bereits oben. Für die Interpretation der PIT-Rohscores werden auf gesonderten Auswertungsblättern geschlechtsspezifische *Grobnormen* (Quartile) angeboten.

**Anwendungsbereiche:** Der PIT eignet sich im Rahmen schuldiagnostischer Zwecke besonders für Gymnasiasten und Realschüler ab der 9./10. Klasse zur Erfassung bildungs- und berufsrelevanter Interessen.

**Kritik:** Im Vergleich zur (einfachen) Durchführung des PIT erscheint die Testauswertung (Auszählung und Gruppierung der Rohscores) etwas umständlich und zeitraubend. Für die Applikation in Fragebogenform (Gruppenverfahren) fertigten wir spezielle Auswertungsschablonen an. Obgleich der PIT ein relativ objektives, zuverlässiges und valides Interessendiagnostikum bietet und auch hinsichtlich einzelner Persönlichkeitsdimensionen (Selbstkritik versus mangelnde Selbstkritik und Extra- versus Introversion) seine schuldianagnostische Brauchbarkeit erwiesen hat, sind fallweise subjektive Verfälschungen von PIT-Ergebnissen durch Pbn möglich; diese können jedoch nach unseren Erfahrungen bei zusätzlicher Kontrolle mit Hilfe der „Lügenskala“ (Skala A) weitgehend eliminiert werden.

## b) Gruppentests bzw. Gruppenverfahren

Die **Frankfurter Wortschatztests** (WST 5-6 bzw. WST 7-8) von H. Anger, R. Bargmann und E. Hylla, neubearbeitet von H. Horn (Beltz-Verlag)

Der WST ist als Begabungstest am Deutschen Institut für Internationale Pädagogische Forschung in Frankfurt/Main entwickelt worden. Er liegt in zwei Versionen vor: WST 5-6 (für 5. und 6. Klassen) und WST 7-8 (für 7. und 8. Klassen).

**Testaufbau:** Es werden sog. Schlüsselwörter (insgesamt 60 Testitems) vorgegeben, denen jeweils aus 5 Wahlantworten die *eine* richtige (Lösungswort) zuzuordnen ist. Zwei Beispielaufgaben mögen das Vorgehen verdeutlichen:

<i>beschwichtigen</i>	<i>Frist</i>
A ..... beschweren	A ..... Dachkante
B ..... verschweigen	B ..... Mahnung
C ..... beschönigen	C ..... Futtermenge
D... besänftigen	D ..... Zeitraum
E...schwächen	E... Datum

Sowohl für den WST 5-6 als auch für den WST 7-8 liegen Parallelförmungen vor.

**Erfassungsdimensionen:** Nach Anger & Bargmann (1954, S. 4) prüft der WST den Wort- und Begriffsschatz, sprachliches bzw. sprachlogisches Denken und Verständnis, Analogiedenken u.ä. Somit werden die Faktoren *Verbal Comprehension* und *Verbal Reasoning* erfaßt (vgl. Heller 1970a, S. 100 f.). Ferner sprechen die Testautoren dem WST die Funktion eines *allgemeinen* Begabungstests zu.

**Testgütekriterien:** Die Paralleltest-*Reliabilität* liegt zwischen 0.92 und 0.93 (Anger & Bargmann 1954, S. 22 f.). Die *Objektivität* des WST ist gleichfalls gut. Zur *Validität* des WST liegen unterschiedliche Angaben vor; je nach Test- bzw. Außenkriterium wurden Übereinstimmungskoeffizienten zwischen 0.33 (Schulbewährung nach 1 Jahr auf Realschule und Gymnasium) und 0.54 (Schulerfolg nach 4 Jahren) bzw. von 0.59 (WST 5-6 in Relation zum Frankfurter Analogietest FAT 5-6) und 0.45 (WST 5-6 in Relation zum Begabungstestsystem BTS von W. Horn) errechnet. Als *Normen* werden Prozentrangplätze bzw. PR-Normenbänder (s.S. 70f. oben) angeboten, differenziert nach Schultyp und Schulortsgroße bzw. voll- versus weniger gegliederten Volksschulen. Spezielle WST-Normen für *taube/resthörige* Pbn (vgl. Tab. IVa) und *schwerhörige* Pbn (vgl. Tab. IVb) sowie für hörgeschädigte Real- und Wirtschaftsschüler (vgl. Tab. IVc) finden sich im Anhang dieses Buches. Entsprechende WST-Normen für Sehgeschädigte (*Blinde* und *Sehbehinderte*) enthalten die Veröffentlichungen von Heller & Köhn (1973) sowie Heller & Schirmer (1973).

**Anwendungsbereiche:** Die WST 5-6 und WST 7-8 werden im 5. bis 8. Schuljahr zur allgemeinen (verbalen) Begabungsdiagnostik resp. zur Schuleignungsprognostik eingesetzt (z.B. Heller 1970a, 1970b). Darüberhinaus dienen Wortschatztests dieser Art der Prüfung des lexikalischen Bestandes, des sprachlogischen Sinnverständnisses u.ä.

Im sonderpädagogischen Bereich wurde der WST von uns versuchsweise bei *Hör-/Sprachgeschädigten* eingesetzt (vgl. Kap. III, 2b unten); die so ermittelten – vorläufigen – Normdaten sind in Tab. IV beigegeben. Bei *blinden* und *sehbehinderten* Schülern führten neuer-

dings Köhn (1970) und Schirmer (1971) umfangreiche WST-Erhebungen durch. Die zu diesem Zweck in (Blinden-)Punktschrift und vergrößerten Schwarzdruck adaptierten Originalformen wurden zusammen mit den getrennt ermittelten WST-Normen für Sehgeschädigte Anfang 1973 publiziert (Beltz) und stehen somit Interessenten zur Verfügung. Im entsprechenden Testmanual finden sich auch Angaben über die Testgütekriterien (Retest reliabilitätskoeffizienten zwischen 0.71 und 0.97 (Blinde) sowie 0.78 und 0.94 (Sehbehinderte), jeweils nach 21 Monaten gültig) bzw. zur Validität.

**Testkritik:** Der WST ist ein objektives, zuverlässiges und zugleich ökonomisches Testverfahren zur Erfassung *verbaler* Fähigkeiten. Als solches ist der Test in seinen Ergebnissen mehr als andere Intelligenztests (z.B. LPS oder PSB) sozio-kulturellen Einflüssen ausgesetzt. Andererseits eignen ihm gerade deshalb schulrelevante Begabungskomponenten (Anger et al. 1965), wenngleich vor übertriebenen Hoffnungen im Hinblick auf die prognostische Gültigkeit, etwa bei Schuleignungsermittlungen, gewarnt werden soll (vgl. Heller 1970a, 1972). Im sonderpädagogischen Kontext hat sich der WST vor allem bei Intelligenzuntersuchungen *blinder* und *sehbehinderter* Hauptschüler bewährt. Hingegen ist seine Anwendung bei hörgeschädigten resp. sprachlich (soziokulturell versus sensorisch) deprivierten Jugendlichen nicht unproblematisch (vgl. Eckrich 1970).

**Fremdsprachen-Eignungstest für die Unterstufe (FTU 4-6)** von W. Correll und K. Ingenkamp bearbeitet (= deutsche Parallelentwicklung zur Elementarform des „Modern Language Aptitude Tests“ von John B. Carroll & Stanley M. Sapon) (Beltz Verlag).

**Testaufbau:** Der FTU 4-6 besteht aus 5 Teilen mit 89 Aufgaben, von denen jedoch nur 75 Items gewertet werden; die erreichbare Höchstpunktzahl beträgt 100 RP. Die 14 nicht gewerteten Zusatzitems sollen lediglich die schnelleren Schüler beschäftigen. Die gesamte Durchführung des FTU dauert 75 Minuten. Folgende Subtestreihen werden angeboten (zur Prognose dient aber die *Gesamtleistung* im FTU):

- I Lückenwörter
- II Passende Wörter
- III Reimwörter
- IV Zahlenlernen
- V Phonetische Wörter

Aufgaben und Anweisungen zum FTU sind auf Tonband aufgenommen und werden zusammen mit dem Testmanual (Auswertungsrichtlinien, Normentabellen u.ä.) und den Arbeitsheften vom Verlag geliefert.

**Erfassungsdimensionen:** Der FTU 4-6, der keinerlei Fremdsprachenkenntnisse voraussetzt, soll „Aussagen über die individuelle Eignung für den Fremdsprachenunterricht“ ermöglichen. „Mit seiner Hilfe kann die Gültigkeit prognostischer Aussagen über die Erfolgsaussichten beim Fremdsprachenlernen wesentlich verbessert werden“ (Correll & Ingenkamp 1967, S. 3).

Faktorenanalytische Untersuchungen liegen bisher nur zur Originalform des FTU, d.h. der Elementarform der Modern Language Aptitude Tests (MLAT) vor. „Die erfaßten Faktoren der Sprachfähigkeit waren vor allem (nach den Bezeichnungen von French): 1. verbal knowledge factor (V) (Wortkenntnis), 2. word-fluency factor (W) (Wortflüssigkeit), 3. Fluency-of-expression factor (FE) (Ausdrucksflüssigkeit), 4. associative memory factor (M) (Assoziatives Gedächtnis), 5. naming factor (Na) (Benennung)“ (Correll & Ingenkamp 1967, S. 11). Aufgrund der Untersuchungsergebnisse mit verschiedenen Testvorformen wurden schließlich 4 Dimensionen der Fremdspracheneignung von den amerikanischen Autoren formuliert, die dann als Konstruktionsgrundlage der MLAT bzw. des FTU dienen: „1. *Phonetic coding*“, das ist die Fähigkeit, auditiv dargebotenes phonetisches Material so zu entschlüsseln, daß es wiedererkannt, identifiziert und einige Zeit lang behalten werden kann. Diese Fähigkeit wird für sehr wichtig gehalten, und ihre geringe Ausprägung soll zu Schwierigkeiten beim Erinnern phonetischen Materials und bei der Nachahmung von Aussprache

und Intonation führen. Im FTU 4-6 wird diese Fähigkeit vor allem durch den Teil 'Phonetische Wörter', wahrscheinlich aber auch durch die Teile 'Zahlenlernen' und 'Reimwörter' angesprochen. 2. '*Grammatical Sensitivity*', das ist die Fähigkeit, die Funktion von Wörtern in vielfachen Zusammenhängen zu erkennen und Formen und Anordnungen so zu handhaben, wie sie dem Sprachgebrauche angemessen sind. Diese Fähigkeit wird im Teil 'Passende Wörter' des FTU 4-6 erfaßt. 3. '*Rote Memory for Foreign Language Materials*', das ist die Fähigkeit, Fremdsprachenmaterial zuverlässig in relativ kurzer Zeit zu speichern. Im MLAT, Form A und im FTO 9+ ist der Teil 'Wortpaare' zur Untersuchung dieser Fähigkeit vorgesehen. Im FTU 4-6 wird diese Fähigkeit im Teil 'Zahlenlernen' erfaßt. 4. '*Inductive Language Learning Ability*', das ist die Fähigkeit, sprachliche Formen, Regeln und Strukturen aus neuem sprachlichen Zusammenhang mit einem Minimum an Anleitung und Überwachung selbst zu erschließen. Diese Fähigkeit konnte im MLAT noch nicht in befriedigender Weise erfaßt werden" (loc. cit; kursiv v. Verf.).

**Testgütekriterien:** Tonband-Instruktion und Auswerteschablonen sollen weitgehende *Objektivität* des FTU 4-6 gewährleisten. Die Halbwerts-*Reliabilität* beträgt  $r_{tt} = 0.96$  (nach Spearman-Brown aufgewertet); der *Standardmeßfehler* liegt bei 4 RP. Zur *Validität* des FTU 4-6 werden im Testmanual diverse Untersuchungsergebnisse (unterschiedlicher Provenienz) mitgeteilt. So korrelieren etwa HAWIK und FTU 4-6 mit  $r = 0.62$ , FTU 4-6 und ELT 6-7 (Englisch-Leistungs-Test von J. Kamratowski) mit  $r = 0.51$  oder FTU 4-6 und B 6 (Berliner Englisch-Leistungstest) mit  $r = 0.56$ . Zu den standardisierten Englischzensuren werden Zusammenhänge von  $r = 0.35$  bis  $r = 0.69$  oder korrigierte Validitätskoeffizienten um 0,8, zu den standardisierten Deutschzensuren r-Koeffizienten um 0,6 angegeben. Zwischen FTU 4-6 und einigen Gruppen-Intelligenztests (FAT 4-6 und BTS) sollen Korrelationen zwischen 0,7 und 0,8 bestehen (loc. cit., S. 13 ff.). Beck (1969, S. 79), die unter Anleitung d. Verf. verschiedene Tests und Schulleistungszensuren auf ihre Prädiktorfunktion hin in bezug auf den Englischerfolg in der Hauptschule untersuchte, konnte die entsprechenden im Testmanual aufgeführten Werte (Rangkorrelationen zwischen 0,17 und 0,70, im Median  $\rho = 0.47$ ) im großen und ganzen bestätigen. Allerdings korrelierten bei der Beckschen Untersuchung PSB sowie die Zensuren in Deutsch, Rechtschreiben und Lesen insgesamt höher (zwischen  $r = 0.7$  und  $r = 0.5$ ) mit der – späteren – Englischzensur als FTU 4-6 und Englischzensur ( $r = 0.45$ ); siehe dazu ausführlicher Kap. III, 1 unten. Nach einer Untersuchung von Schollmeyer (1970), der den FTU 4-6 bei schwerhörigen Realschülern erprobte, korreliert der FTU 4-6 mit der späteren Englischzensur mit  $r = 0.61$ , mit dem Hörverlust (mittl. dB) immerhin noch um 0,4, wohingegen zwischen HAWIK-Handlungsteil und FTU bzw. HAWIK-Handlungsteil und Englischzensur versus Hörverlust praktisch keine Zusammenhänge nachzuweisen waren ( $r = 0.0$  bzw. 0,1 versus 0,2); vgl. auch Kap. III, 2b unten. Die *Klassen-Normen* zum FTU 4-6 werden in T-Werten bzw. T-Wert- und %-Rang-Bändern (s.S. 70f.) angegeben und reichen von der 4. bis zur 6. Klasse.

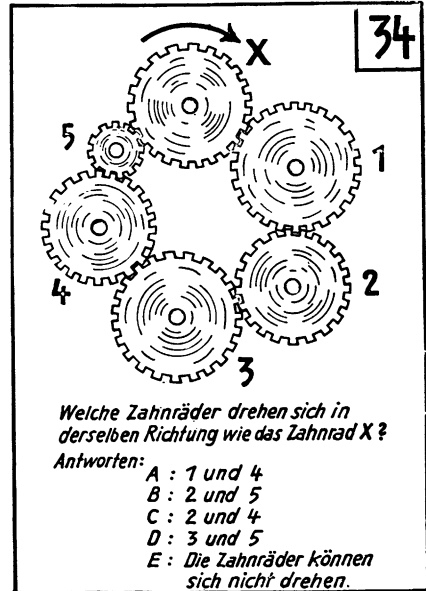
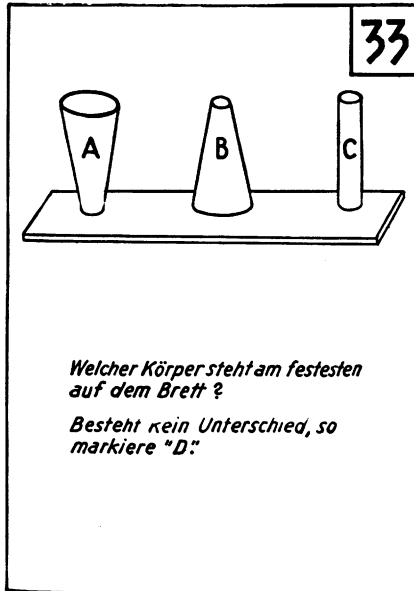
**Anwendungsbereiche:** Der FTU 4-6 empfiehlt sich zur Ermittlung der Fremdspracheneignung (besonders für Englisch) in den Schuljahren 4 bis 6, d.h. vor Beginn des Fremdsprachenunterrichts. Für Anwendungszwecke im sonderpädagogischen Bereich, z.B. bei Schwerhörigen, Sprachbehinderten, aber auch bei Sehbehinderten und Blinden, werden weitere Untersuchungen sowie gewisse Modifikationen der Testdarbietung (teilweise) unumgänglich sein.

**Testkritik:** Obwohl der FTU 4-6 nach Aussage der Testautoren sowie einer Reihe vorliegender Testgütekriterien für die Begabungsermittlung, insonderheit die Erfassung der Fremdspracheneignung, ein brauchbares Diagnostikum darstellt, muß doch kritisch angemerkt werden, daß offensichtlich die Prädiktorfunktion bestimmter Schulleistungszensuren (Deutsch bzw. Rechtschreiben und Lesen) oder gut standardisierter Begabungstests (z.B. PSB) in diesem Punkte dem FTU in nichts nachsteht. Somit wird sich in vielen Fällen die Frage der Testökonomie, d.h. eines möglichen FTU-Verzichts zugunsten einer differenzierteren Erfassungsstruktur mit Hilfe anderer Tests, stellen. Trotz-

dem sollte der FTU 4-6 hinsichtlich seiner eigentlichen Vorhersagegültigkeit bei Grund- und Hauptschülern resp. Teilen der Sonderschule weiterhin kontrolliert bzw. verbessert werden, zumal kein vergleichbares Testinstrument derzeit auf dem deutschen Markt ist. Ferner bedarf die sozio-kulturelle Determination der FTU-Leistung präzisierter Angaben.

### Mechanisch-Technischer Verständnistest (MTVT) von G.A. Lienert (Verlag Dr. Hogrefe)

**Testaufbau:** Der MTVT besteht aus 32 praktisch-technischen Problemaufgaben, z.B. Hebel-aufgaben, Räderwerkitems u.ä. (vgl. die abgebildeten Übungsbeispiele Nr. 33 und 34). Trotz der – aus ökonomischen Gründen – vorgesehenen zeitlichen Limitierung auf 45 Minuten ist der MTVT ein Power- oder Niveautest.



**Erfassungsdimensionen:** Der MTVT soll das mechanisch-technische Verständnis oder Denken prüfen, das nach Lienert einen wichtigen Aspekt der „praktischen Intelligenz“ repräsentiert. Die Faktorenanalyse mit einer Testvorform erbrachte folgendes Ergebnis: „Der MTVT ist durch einen Faktor C, interpretiert als die Fähigkeit des zweckmäßigen Kombinierens und sinnvollen Ergänzens, zu 67 % ( $r = 0.82$ ) bestimmt. Ein weiterer Faktor A, interpretiert als die Fähigkeit des schlußfolgernden Denkens, ist mit einem Anteil von 33 % ( $r = 0.57$ ) an der MTVT-Leistung beteiligt. Ein dritter Faktor B, interpretiert analog dem kinaesthetischen Faktor von Thurstone, ist in der MTVT-Leistung nicht wirksam“ (Lienert 1964, S. 7).

**Testgütekriterien:** Der MTVT verfügt über optimale *Objektivität*. Die *Reliabilität* läßt indessen manche Wünsche offen. So wurden vom Testautor Retest-Zuverlässigkeitskoeffizienten von  $r = 0.76$  bzw. (korrigiert)  $0.82$  ermittelt, wobei der Übungsgewinn (der Wiederholungstestung nach 7 Wochen) bei durchschnittlich 6 Z-Werten lag. Die Halbierungszuverlässigkeit des MTVT beträgt  $r = 0.81$  bzw.  $0.86$ , die innere Konsistenz  $r = 0.84$ . Der *Standardmeßfehler* liegt bei 5 Z-Werten.

Abgesehen von der faktoriellen Gültigkeit des MTVT (siehe oben) liegen Informationen zur inneren und äußeren *Validität* vor. Die Interkorrelationen des MTVT zu verschiedenen Subtests der Mechanical Aptitude Battery von Thurstone variieren zwischen  $0.24$  und  $0.43$ , zu verschiedenen Subtests des LPS zwischen  $0.32$  und  $0.40$  (Lienert 1964, S. 7 f.). Wir selbst



ermittelten in der Relation MTVT und LPS 7–10 folgende Interkorrelationen:  $r = 0.28$  (bei 881 Gymnasiasten und 523 Realschülern) sowie  $r = 0.30$  (bei 2664 Hauptschülern); die von Lienert bei 209 Oberschülern gefundenen Zusammenhangswerte von MTVT und LPS konnten somit nicht ganz bestätigt werden (vgl. Heller 1970a, S. 103). Mit dem FLT (Form-Lege-Test von G. A. Lienert) ist die MTVT-Leistung mit  $r = 0.44$  korreliert. Berücksichtigt man, daß fast alle (verfügbaren) „Technischen“ Intelligenztests relativ niedrige Reliabilitäts- und Validitätskoeffizienten aufweisen, dann kann man der Meinung des Testautors beipflichten, wonach der MTVT ein „objektiver, im Verhältnis zu den meisten Tests dieser Art hoch zuverlässiger Test“ ist, dessen Gültigkeit „inhaltlich und faktoriell begründet“ wird.

Die *Normen* werden in PR- und Standardwert- oder Z-Einheiten ( $M = 100$ ,  $s = 10$ ) angegeben. Es liegen Schultypen-, Berufsgruppen- und Altersgruppen-Normen vor.

**Anwendungsbereiche:** Der MTVT dient der Erfassung mehr praktisch-technischer Fähigkeiten und ist ab dem 13. Lebensjahr auch im Sekundarschulwesen einsetzbar. Nach unseren Erfahrungen stellt er eine gute Ergänzung zum LPS 7–10, besonders bei Haupt- und Realschülern, dar. Für die interschulische Begabungsdifferenzierung im traditionellen Sekundarschulsystem (Hauptschule, Realschule, Gymnasium) ist er hingegen weniger geeignet (Heller 1970a, 1970b).

**Testkritik:** Wie andere Technische Intelligenztests kann auch der MTVT hinsichtlich seiner Treffsicherheit nicht allen Ansprüchen moderner Testkritik genügen. Auf der Hauptschulabschlußstufe sowie in der Eignungsermittlung von Berufsschülern (Lehrlingen und Anlernlingen) ist der Test jedoch insgesamt ein recht brauchbares Hilfsmittel zur Information über die praktisch-technische Begabung von Schülern und Jugendlichen. Auch im sonderpädagogischen Bereich (z.B. bei Hör-/Sprachgeschädigten, möglicherweise auch bei Lernbehinderten sowie Körperbehinderten) erscheint die Anwendung des MTVT lohnend.

### Der Figuren-Durchstreich-Test nach B. Bourdon oder Bourdon-Test (TZ)

Die sog. Durchstreich-Tests „gehen auf Anregungen von B. Bourdon (1895, 1902) zurück, der seinen Probanden eine Textseite z.B. aus einem Buch über Säugetiere vorlegte und alle a, r, s, i mit einem Bleistift durchstreichen ließ (1902). Er beachtete vor allem die pro Blatt benötigte Zeit, aber auch Fehler in Form von Auslassungen und Fehldurchstreichungen“ (Bartenwerfer 1964, S. 390). Der Bourdon-Test erfuhr mittlerweile eine Reihe von Veränderungen bzw. Abwandlungen. Hier soll eine Form beschrieben werden, die statt Buchstaben eine Folge von einander ähnlichen *Figuren* enthält (siehe nachstehende Abbildung). Bezüglich der Standardversuchsbedingungen verweisen wir auf Martin (1961).



**Testaufbau:** Der Figuren-Durchstreich-Test nach Bourdon besteht aus einem Testblatt im DIN A 4-Format, das insgesamt 750 Figuren enthält. Aus den Figuren sind folgende vier Typen herauszufinden und – sooft sie vorkommen – durchzustreichen. Ein Testblatt enthält jeweils 159 Lösungsfiguren. Die Standardversuchsbedingung sieht eine Bearbeitungszeit von 10 Minuten vor; der Bourdon-Test ist ein reiner Speedtest.



**Erfassungsdimensionen:** Mit dem Bourdon-Test soll die sog. Tenazität, d.h. Zähigkeit oder Hartnäckigkeit der Aufmerksamkeit oder bei der Konzentration, geprüft werden.

**Testgütekriterien:** Bei Einhaltung bestimmter Standardprüfbedingungen, die jedoch in verschiedenen Varianten vorliegen, ist die *Objektivität* des Bourdon-Tests ziemlich problematisch. Die Vergleichbarkeit der Versuchsbedingungen resp. deren Kontrolle ist freilich unerlässlich, wenn es um die Erfassung interindividueller Differenzen – hier in bezug auf die Konzentrationsfähigkeit – geht. Daraus folgt, daß auch die Bestimmung der *Reliabilität* des Bourdon-Tests jeweils in engem Zusammenhang zur Standardprüfsituation steht. M.a.W.: Der Bourdon-Test ist nur dann als „Test“ anzusprechen, wenn er den Testgütekriterien genügt resp. diese unter konkretem Bezug auf bestimmte Standardprüfmuster, also dieser oder jener zu beschreibenden Testform, kontrolliert und für ausreichend befunden wurden. Entsprechend werden in der Literatur sehr unterschiedliche Zuverlässigkeitsaussagen gemacht. Zumeist auf Testhalbierungen basierende Angaben variieren zwischen  $r_{tt} = 0.83$  und  $r_{tt} = 0.96$  (korrigierte Werte). Noch weniger befriedigen die Angaben zur *Validität* des Bourdon-Tests, die je nach Außenkriterium Werte um 0.4 selten überschreitet; gelegentlich werden aber auch höhere Validitätskoeffizienten (etwa in der Relation von Bourdon-Test und Schulzensuren versus Lehrerurteilen) genannt. *Normen* zum Bourdon-Test liegen nicht vor. Da der Figuren-Durchstreich-Test vielfach noch heute in der testdiagnostischen Praxis Verwendung findet, z.B. bei Hör- und Sprachgeschädigten, aber auch in der Regelschule, der klinischen Diagnostik oder bei der Berufsberatung, haben wir an Hand des von E. Martin (1961) erstellten Datenmaterials Alters-Normen für taube und resthörige Schüler berechnet und im Appendix hier beigelegt (vgl. Tab. Va und Vb).

**Anwendungsbereiche:** Der Bourdon-Test kommt eigentlich heute nur noch für experimentalspsychologische Untersuchungszwecke im Rahmen der Erfassung von Aufmerksamkeit, Konzentration u.ä. Hilfs- oder Stützfunktionen der Intelligenz in Frage. Auch mag er für bestimmte klinische und/oder sonderpädagogische Schulgruppen, z.B. Hör- und Sprachgeschädigte, Legastheniker etc., nach wie vor interessant sein. Für entsprechende Anwendungszwecke müßten aber mindestens genaue Instruktions- und Durchführungsbedingungen mitgeteilt sowie detaillierte Auswertungsrichtlinien und Normmaßstäbe vorgelegt werden.

**Testkritik:** Der Bourdon-Test kann nur bedingt als „Test“ angesprochen werden, da wichtige Elemente einer Standardisierung bislang fehlen. Zwar steht nach Hiltmann (1969, S. 147) „die praktische Bewährung des im *Bourdon-Test* verwirklichten methodischen Prinzips außer Zweifel. Indessen sind neuere experimentelle, klinische oder statistische Brauchbarkeitskontrollen nicht bekannt.“ Auch die von uns mitgeteilten (Grob-)Normen zum Bourdon-Test mildern die berechtigte Testkritik nur graduell und zudem nur partiell im Hinblick auf bestimmte, mehr oder weniger eng umschriebene Pbn-Gruppen. Vielfach wird man bei der Ermittlung der Konzentrationsfähigkeit und anderer Eigenschaften der sog. Sekundärintelligenz auf brauchbarere Verfahren (z.B. KLT oder Pauli-Test) zurückgreifen können.

**Der Konzentrations-Leistungs-Test (KLT) von H. Düker & G. A. Lienert (Verlag Dr. Hogrefe)**

**Testaufbau:** Beim KLT handelt es sich um einen Rechentest mit verschiedenen Schwierigkeitsversionen, nämlich Schwierigkeitsstufe C (nur für Volksschüler bzw. Hauptschüler) und Schwierigkeitsstufe D (für Berufs- und Oberschüler resp. Erwachsene). Zwei Übungsbeispiele der Schw.St.D mögen das Vorgehen erläutern:

$$\begin{array}{ll} \text{Beispiel A: } 8 + 9 - 2 & \text{Beispiel B: } 3 + 6 - 8 \\ & 5 - 4 + 3 \qquad \qquad 9 + 1 + 7 \end{array}$$

Zunächst müssen immer beide Zeilen einer Aufgabe ausgerechnet und die Ergebnisse gemerkt (im Kopf behalten) werden. Sodann werden die beiden Ergebnisse unter Beachtung folgender zwei Regeln miteinander verglichen: Sofern das Ergebnis der ersten (oberen) Zeile *größer* ist als das der zweiten (unteren) Zeile, müssen die Pbn die untere Zahl, also das zweite Ergebnis, von der oberen Zahl *abziehen* und das Endergebnis in das Kästchen eintragen (in Beispiel A wäre 11 einzutragen). Sofern das Ergebnis der ersten Zeile *kleiner* ist als das der zweiten Zeile, sind beide Resultate zu *addieren* und das Endergebnis wieder in das leere Kästchen einzutragen (18 in Beispiel B). Insgesamt sind 250 Items dieser Art in 30 Minuten zu bearbeiten.

**Erfassungsdimensionen:** Mit Hilfe des KLT sollen Aufmerksamkeit und Konzentration erfasst werden. Daneben sind wohl auch wichtige Eigenschaften des Arbeitsverhaltens, z.B. Ausdauer, Anstrengungsbereitschaft, Arbeitssorgfalt, Belastbarkeit u.ä. an der KLT-Leistung beteiligt.

**Testgütekriterien:** Die *Objektivität* der Testdurchführung und -auswertung ist beim KLT gewährleistet. Die *Reliabilität* der KLT-Mengenleistung ist gut; die Stabilitätskoeffizienten liegen bei sofortiger Retestung nach 1 Stunde bei 0.92, bei Retestung nach 8 Wochen zwischen 0.88 und 0.86. Die entsprechenden Retestreliabilitäten in bezug auf den Fehlerprozensatz im KLT sind dagegen weniger gut; die Koeffizienten liegen zwischen 0.74 und 0.05 nach 8wöchiger Retestung bzw. bei 0.47 nach sofortiger (1stündiger) Retestung. Allerdings ist der Übungsgewinn bei sofortiger Wiederholung des KLT beträchtlich, er beträgt zunächst (Retestung nach 1 Stunde) rd. 25 % und sinkt dann allmählich, bis er nach 2 Jahren praktisch völlig verschwunden ist (Angaben der Testautoren). Der *Standardmeßfehler* wird mit  $s_e = 3,3 Z$  (für die KLT-Mengenleistung) bzw.  $s_e = 1/2 F \%$  (für die Arbeitsqualität) angegeben.

Über die *Validität* des KLT ist – wie bei den meisten Tests dieser Art – wenig bekannt. Im allgemeinen korreliert die KLT-Leistung mit den Schulzensuren mäßig bis niedrig, am höchsten (0.4) mit der Mathematikzensur (besonders bei den jüngeren Pbn des 4. bis 6. Schuljahrs). Auch zum IST von Amthauer wurden keine engeren Beziehungen der KLT-Leistung ermittelt. Die auf Lehrerurteil basierenden Gültigkeitskoeffizienten liegen maximal bei  $r = 0.50$ , durchschnittlich jedoch nur bei 0.25; bezüglich des F % wurden überwiegend negative Korrelationskoeffizienten gefunden. Wir selbst ermittelten bei 220 Gymnasiasten und Realschülern der 10. und 12. Klasse in der Relation von KLT-Leistung und LPS-APR 1+ (4 Min.) relativ hohe Korrelationskoeffizienten zwischen 0.79 und 0.49 oder im Mittel  $r = 0.69$ ; in der Relation von KLT-Leistung und LPS-APR 1+2+ (8 Min.) errechneten wir sogar r-Werte zwischen 0.81 und 0.51 oder im Mittel  $r = 0.74$  (vgl. Heller 1970a, S. 102 f.).

Zum KLT werden diverse PR- und Standardwert- oder *Z-Normen* ( $M = 100$ ,  $s = 10$ ) angeboten, und zwar Altersnormen für männliche versus weibliche Berufs- und Berufsfachschüler, Klassennormen für männliche versus weibliche Volks- und Oberschüler sowie diverse Berufsgruppen-Normen für männliche versus weibliche Lehrlinge und Anlernlinge. Für Mittel- bzw. Realschüler stehen keine Normen im Testmanual. Anhand einer Realschulstichprobe der 10. Klasse ( $N = 52$ ) errechneten wir deshalb folgende Statistiken:  $M = 110$  RP und  $s = 40$  RP für die KLT-Mengenleistung sowie  $M = 11$  Fehler für die KLT-Fehlervariable. Die beobachteten Differenzen bezüglich der KLT-Leistung und KLT-Fehlerprozente zwischen Jungen und Mädchen waren statistisch nicht zu sichern (loc. cit.).

**Anwendungsbereiche:** Der KLT ist ab der 4./5. Klasse zur Erfassung der „allgemeinen psychischen Leistungsfähigkeit“, also der psychischen Aktivität oder „Willensstärke“, der Anstrengungsbereitschaft und Leistungsmotivation, der Konzentration, Ausdauer u.ä. Eigenschaften der sog. Sekundärintelligenz, geeignet. Sein Einsatz dürfte ohne größere Mühe auch in der Sonderschule, etwa bei Hör- und Sprachgeschädigten, möglich und informativ sein, wenngleich bislang hierzu noch keine Normen vorliegen.

**Testkritik:** Während die Reliabilität des Faktors „Leistungsmenge“ im KLT als gut bezeichnet werden kann, schwankt die Zuverlässigkeit der Fehler-Variablen teilweise beträchtlich und genügt somit nicht immer den teststatistischen Anforderungen. Als Indikator der Konzentrationsleistung ist der KLT ziemlich ohne Einschränkung zu empfehlen, wohingegen diagnostische Aussagen über die Arbeitsqualität nur unter Vorbehalt möglich sind. Hierzu bedarf es zusätzlicher Informationsquellen, etwa via Verhaltensbeobachtung oder Exploration resp. Anamnese. Siehe auch Bartenwerfer (1964).

**Der Pauli-Test.** Anweisung zur sachgemäßen Durchführung, Auswertung und Anwendung des Kraepelinschen Arbeitsversuches nach R. Pauli, herausgegeben von W. Arnold (Verlag J. A. Barth, München, oder TZ).

**Testaufbau:** Der Pauli-Test – eine 1936 von R. Pauli erstmals in standardisierter Form vorgestellte Variante des Kraepelinschen Arbeitsversuchs aus den neunziger Jahren des vorigen Jahrhunderts – besteht aus einem Bogen voller einstelliger Zahlen, die kolonnenförmig angeordnet sind. Jeweils zwei übereinanderstehende Zahlen sind fortlaufend zu addieren und die letzte Stelle des Summenergebnisses zwischen die beiden Zahlen zu schreiben. Der Pauli-Test ist ein reiner Speedtest und dauert bei voller Durchführung 60 Minuten; alle 3 Minuten wird vom Pb auf ein Glockenzeichen des Testleiters hin die zuletzt bearbeitete Addition markiert. Für die Auswertung und Interpretation interessieren somit u.a. die Gesamtzahl der Additionen, die Fehlerzahl, der Arbeitsverlauf (3 Minuten-Leistungseinheiten) u.ä. Die neuesten Informationen über systematische Untersuchungen zur Standardisierung des Pauli-Tests finden sich in der Monographie von W. Arnold.

**Erfassungsdimensionen:** Mit Hilfe des Pauli-Tests werden praktisch wieder dieselben Merkmale wie beim KLT und ähnlichen Verfahren erfaßt, also Konzentration, Belastbarkeit, Ausdauer, Leistungsmotivation, Arbeitssorgfalt u.ä. Nach einer faktorenanalytischen Untersuchung zum Pauli-Test (unter dem Aspekt der „Leistungsmenge“) von G. Bäumler (1964) wird vor allem die „Grundfähigkeit des rechnerischen Umgangs mit Zahlen“ im Pauli-Test geprüft.

**Testgütekriterien:** Die *Objektivität* der Testdurchführung und Testauswertung ist gesichert. Die Angaben zur *Reliabilität* des Pauli-Tests schwanken, zumal auch hier wieder ein mehr oder minder großer Übungs- und Lerngewinn veranschlagt werden muß. So fand Plössl 1941 (zit. nach Pawlik 1970, S. 196) „für einen Index aus Menge und Güte eine Retest-Korrelation von 0.96 (Abstand ein Monat). Für andere Indices ist die Zuverlässigkeit geringer. Z.B. haben Bartenwerfer und Plössl für die Fehlerzahl Koeffizienten von 0.7 bis 0.8 gefunden.“ Noch dürftiger sind die Angaben zur *Validität* des Pauli-Tests (vgl. Bartenwerfer 1964, S. 388 ff.). Dies ist um so erstaunlicher, als gerade der Pauli-Test zu den meisteingesetzten Verfahren dieser Art gehört und sich wohl auch praktisch vielfach bewährt hat (vgl. Arnold 1961). Selbst Bartenwerfer meint: „Die verbreitete Wertschätzung des Pauli-Tests kann aus seiner häufigen und erfolgreichen Anwendung auch in der jüngsten Zeit abgeleitet werden. Er fand z.B. Verwendung in der Schulpsychologie (J. Hitpaß 1961), in der Arbeits- und Betriebspsychologie ..., bei der Untersuchung von Studenten ..., bei taubstummen Jugendlichen und für die Eichung eines anderen Tests ...“ (loc. cit.).

Günstiger als die teststatistischen Angaben zu den Gütekriterien der Reliabilität und Validität stellt sich hier das *Normen*-Problem dar. Die Monographie von Arnold enthält beispielsweise rd. zwei Dutzend Pauli-Testnormen, teilweise auf unterschiedliche Untersuchungsbedingungen bezogen. In der diagnostischen Praxis wird man trotzdem sehr oft auf die Erstellung populationspezifischer Normen angewiesen sein, worauf u.a. auch Bartenwerfer hinweist. Für die Belange Hörgeschädigter fügen wir im Anhang dieses Buches eigene resp. von Gruhnwald & Ulich erstellte Pauli-Testnormen bei (vgl. Tab. VIIa bis VIIc).

**Anwendungsbereiche:** Der Pauli-Test ist ab dem 7./8. Lebensjahr anwendbar. Er eignet sich vor allem zur Erfassung der Arbeitsgeschwindigkeit, Anstrengungsbereitschaft, Leistungsmotivation u.ä., aber auch zur Aufhellung der Leistungsdynamik, Arbeitsqualität, Sorgfaltsstreben u.dgl.m. Der Pauli-Test gehört zu den brauchbarsten Verfahren zur Erfassung von Aufmerksamkeit und Konzentrationsleistung und sollte weit mehr noch als bisher gerade auch im schulischen und pädagogischen Bereich bei der Diagnose versus Prognose wichtiger Faktoren der sog. Sekundärintelligenz Verwendung finden.

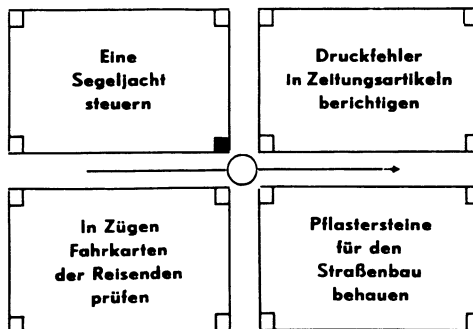
**Testkritik:** Gegenüber vergleichbaren anderen Verfahren bietet der Pauli-Test u.a. den Vorteil, detaillierte *Verlaufsanalysen* wichtiger quantitativer und qualitativer Eigenschaften des Arbeits-Leistungsverhaltens bereits vom Grundschulalter an zu ermöglichen. Relativ ausgedehnte Verhaltensstichproben (Bearbeitungszeit 1 Stunde) garantieren im allgemeinen hinreichend zuverlässige Aussagen über die Arbeitsfähigkeit und Leistungsqualität eines Pb, wenngleich die qualitative Analyse auch hier – wie bei allen vergleichbaren Verfahren – die meisten Probleme in sich birgt. Darüberhinaus besticht der Pauli-Test nicht nur durch seine vielseitigen Verwendungsmöglichkeiten, er ist auch ohne größere testdia-

gnostische Kenntnisse und Erfahrungen seitens des Versuchsleiters leicht und sicher zu administrieren. Seine Auswertung freilich erfordert ein gründliches Studium der schon erwähnten Monographie von Arnold (1961).

Der **Berufs-Interessen-Test (BIT)** von M. Irle (Verlag Dr. Hogrefe)

**Testaufbau:** Der BIT ist in der Form eines vollstandardisierten Fragebogens angelegt und liefert *ipsative* Werte (siehe unten) über verschiedene (Berufs-)Interessenrichtungen. Der Fragebogentest besteht aus zwei Bogen mit je 100 Kästchen oder Testitems, die in 10er Reihen bzw. Kolonnen angeordnet sind, z.B. „Schaufenster dekorieren“, Bankkunden über Vermögensanlage beraten“, „Wetter verschiedener Großstädte vergleichen“, „Buchenwälder auflichten“ (vgl. auch nebenstehende Abbildung). Der Pb soll nun an jedem der

fortlaufend nummerierten (im abgebildeten Beispiel blanc gelassenen) Schnittpunkte („Straßenkreuzung“ eines Viererblocks) die von den vier angebotenen Tätigkeiten jeweils *bevorzugte* markieren, d.h. der Pb wird vor eine *relative* Entscheidung gestellt und muß sich immer für *eine* der vier Wahlmöglichkeiten entscheiden. „Der Wahlzwang bei den einzelnen Items zieht einige Folgen nach sich, die wir betrachten müssen, um die Resultate ... sinnvoll interpretieren zu können. Weil der Proband seine Vorlieben für Tätigkeiten, die verschiedenen Testkategorien angehören, irgendwie in eine Rangordnung bringen muß, kennzeichnen seine Gesamtwerte in erster Linie natürlich nur die relativen Positionen in bezug auf die verschiedenen Interessenvariablen. Solche relativen Angaben von Ausprägungsgraden nennt man *ipsativ* (Cattell 1944). Die Meßwerte gewöhnlicher Art, die nicht aus Items mit Wahlzwang hergeleitet werden, zeigen solche Ausprägungsgrade an, die sich um den Mittelwert der allgemeinen Population verteilen. Sie heißen daher *normative* Werte .... Weiter folgt aus dem Wahlzwang, daß die Durchschnittsergebnisse bei jeder Person etwa gleich sind. Das trifft natürlich nicht den wirklichen Tatbestand der Dinge... Ein Nachteil des ganzen Verfahrens ist der, daß das Profil nicht solche Wesenszüge angemessen wiedergibt, die ausschließlich stark oder ausschließlich schwach sind. Denn zweifellos gibt es Menschen, deren Interessen überwiegend stark ausgeprägt sind, und andere, die durchgehend nur mäßig beruflich interessiert sind. Informationen hierüber sind aber oft sehr wichtig“ (Guilford 1965, S. 203 ff.).



Die BIT-Auswertung geschieht mit Hilfe von Lochschablonen und liefert ipsative Skores zu 9 verschiedenen Berufsinteressenrichtungen (siehe Erfassungsdimensionen).

**Erfassungsdimensionen:** Im BIT werden individuelle Interessenausprägungen in bezug auf folgende Tätigkeitsfelder erfaßt: TH = Technisches Handwerk, GH = Gestaltendes Handwerk, TN = Technische und naturwissenschaftliche Berufe, EH = Ernährungshandwerk, LF = Land- und forstwirtschaftliche Berufe, KB = Kaufmännische Berufe, VB = Verwaltende Berufe, LG = Literarische und geisteswissenschaftliche Berufe, SE = Sozialpflege und Erziehung. Diese 9 Berufsinteressenrichtungen repräsentieren je 18 berufliche Einzeltätigkeiten.

**Testgütekriterien:** Der BIT verfügt sowohl hinsichtlich seiner Durchführung als auch seiner Auswertung über eine optimale *Objektivität*. Die Kontrollen zur internen Itemkonsistenz als Aspekt der *Reliabilität* des BIT erbrachten Koeffizienten zwischen 0.85 und 0.55, im Mittel um 0.87; die Zuverlässigkeitsschätzung nach der Kuder-Richardson-Methode lieferte Werte zwischen 0.94 und 0.75 (vgl. Testmanual zum BIT). Die Interkorrelationen der 9 Interessenrichtungen tendieren leicht zu negativen Werten, was Irle (1955, S. 14) folgender-

maßen interpretiert: „Eine Korrelation von  $r = 0$  oder ein negativer Wert lassen ... nicht den Schluß auf Unabhängigkeit oder negative Abhängigkeit zu, sondern nur auf geringere Abhängigkeit als ein positiver Korrelationswert. Die Tabelle (der hier nicht wiedergegebenen Interkorrelationen; K. H.) gibt ein Bild relativer, aber nicht absoluter Abhängigkeitsgrade der einzelnen Interessenrichtungen voneinander, (was bei entsprechenden anderen Tests oft nicht mitgeteilt wird!).“ Die eigentliche Ursache für die negativen Tendenzen in der Korrelationsmatrix rühren – wie oben bereits angedeutet – von der Anordnung der Testelemente her. „Die Pb haben nur eine positive Wahl zwischen mehreren Alternativen (im PIT besteht weniger Entscheidungszwang, da eine dritte Antwortkategorie „weder-noch“ zur Verfügung steht). Vorteil dieser Art von Testelementen ist, daß die Unterschiede zwischen den verschiedenen Interessen eines Pb klarer zutage treten, Nachteil, daß kein Pb durchwegs hohe oder durchwegs niedrige Skores erhalten kann, bzw. daß schon kleinere Unterschiede zwischen den Interessen innerhalb des Profils künstlich vergrößert werden“ (Mittenecker 1964, S. 476 f.).

Die *Validität* des BIT sieht Irle durch den Nachweis gestützt, „daß die jeweiligen typischen Interessenprofile (des BIT; K. H.) den Berufsbildern und Anforderungen der zugehörigen Berufe entsprechen“ (loc. cit.). Insgesamt werden für 25 verschiedene Berufsgruppen entsprechende BIT-Profile im Testmanual (S. 15) angegeben. Geschlechtsspezifische und bildungsbedingte Einflüsse auf die Interessenausprägungen im BIT konnten Kettel & Simmat (1968) nachweisen. Und Heller (1970a, S. 157 ff. u. S. 193 ff.) erfaßte zahlreiche BIT-Profile von Hauptschülern der 6.-8. Klasse, von Realschülern der 6.-10. Klasse und von Gymnasiasten der 6.-12. Klasse, teilweise nach soziodemographischen Kriterien sowie Schultypen (z.B. altsprachlichen, neusprachlichen und naturwissenschaftlichen Gymnasien) differenziert.

Als *Normen* zum BIT liegen PR für jede der 9 Interessenkategorien vor, so daß für jeden einzelnen Pb ein Profil der Berufsinteressen erstellt werden kann. Niedrige Skores *unter*  $PR = 26$  werden als *schwache* Interessen, hohe Skores *über*  $PR = 75$  werden als *starke* Interessen interpretiert, während mittlere Skores *zwischen dem 26. und 75. PR durchschnittliche* Interessenausprägungen indizieren. Neuerdings stehen auch die von Kettel & Simmat 1968 veröffentlichten geschlechtsspezifischen BIT-Alternormen für Volks- bzw. Hauptschüler, Realschüler und Gymnasiasten (ab dem 13. Lbj.) zur Verfügung. Entsprechende Prozentrangplätze für PH-Studenten berechnete Sayler (1965), wobei zusätzlich die Rangfolge gewählter BIT-Dimensionen berücksichtigt wurde. Nach Mittenecker (loc. cit.) stellt die *Rangordnung* der 9 Interessenscores sogar der wichtigste Auswertungsgesichtspunkt beim BIT dar.

*Anwendungsbereiche:* Der BIT dient der Erkundung beruflicher Interessenrichtungen bei älteren Schülern (in der Regel nicht vor dem 13. Lbj.) sowie Jugendlichen und Erwachsenen. Ursprünglich mehr für die Berufsberatung (z.B. im Rahmen der Berufseignungsfindung) konzipiert, hat der BIT im Laufe der letzten Jahre zahlreiche Erprobungen auch im schulischen Bereich – mit unterschiedlichem Erfolg (vgl. Heller 1970a, 1970b) – erfahren. Über die Ergebnisse unserer BIT-Untersuchungen bei Hör-/Sprachgeschädigten wird später noch zu berichten sein (vgl. Kap. III, 2b).

*Testkritik:* Als „direkte“ Methode vermittelt der BIT nur dann zuverlässige und gültige Informationen, wenn sichergestellt ist, daß der Pb gemäß der Instruktion sorgfältig und ehrlich seine Wahlen trifft und in etwa schon eine Vorstellung über die gefragten Tätigkeiten besitzt. Unter diesen Voraussetzungen eignet sich der BIT besser als etwa der PIT zur Erfassung berufsrelevanter Interessenrichtungen bei *Haupt-* und *Berufsschülern*, wohingegen er hinsichtlich seiner Aufgabenstruktur (Iteminhalte) weniger gut die Bildungs- und Berufsinteressen von Realschülern oder gar Gymnasiasten und Studenten trifft (letztere werden im PIT besser repräsentiert). Diese Kritik schränkt die Anwendung des BIT im Bildungsbereich (allgemeinbildenden Schulwesen), z.B. bei der Schul- und Studieneignungs-ermittlung, ein. Daß er trotzdem auch hier immer wieder zum Einsatz kam, erklärt sich in erster Linie aus dem großen Mangel entsprechend geeigneter bzw. verfügbarer Instrumente zur Interessenerfassung. Bildungs- und Berufsinteressen gehören jedoch zu den wichtigsten

Stützfunktionen der Intelligenz und müssen entsprechend im Intelligenzdiagnostischen Design berücksichtigt werden. Der BIT bietet hierfür – je nach Untersuchungspopulation resp. Interessenlage der Pbn – ein mehr oder minder brauchbares Diagnostikum. In Zukunft werden wir weitere, vor allem schulisch und bildungsrelevantere Interessentests dringend benötigen.

Das Methodenreferat ist verhältnismäßig umfangreich geraten. Es wird jedoch – so hoffe ich – gerade für die Intelligenzdiagnostische Praxis in der Schule und Sonderpädagogik wertvolle Informationen bieten. Aus Gründen der inneren Logik sowie des besseren Verständnisses von Testmethoden und ihrer angemessenen Wertbeurteilung wurden bereits im vorstehenden Kapitel zahlreiche neuere, teilweise selbst erarbeitete oder zusammengetragene, Untersuchungsergebnisse erörtert. Um Redundanz zu vermeiden, werden wir im folgenden hierauf nur noch verweisend eingehen. Wer sich umfassender über psychodiagnostische Tests informieren möchte, sei abschließend nochmals auf Heiß (1964) und Hiltmann (1969) sowie Marsolek & Ingenkamp (1968) verwiesen.

### III. Problembereiche und Befunde intelligenzdiagnostischer Untersuchungen in der Schule und Sonderpädagogik

#### I. Intelligenzdiagnostische Probleme in der Grundschule und den weiterführenden schulischen Bildungseinrichtungen (Hauptschule, Realschule, Gymnasium)

Bei intelligenzdiagnostischen Untersuchungen in der Schule rücken zwei Problemkomplexe in den Vordergrund: a) das Problem der *Sicherung inter- versus intraindividueller Testleistungsdifferenzen* bzw. die Frage der Interpretation von *Profildifferenzen*, b) Probleme der *Schuleignungsermittlung* und damit zusammenhängende Fragen nach den *Schuleignungskriterien*. In der intelligenzdiagnostischen Praxis bestehen freilich zwischen beiden Problemkomplexen mannigfache Interdependenzen. Unsere getrennte Betrachtung erfolgt in erster Linie aus didaktischen und erst in zweiter Linie aus sachimmanen Gründen.

Ad a) Die Interpretation von Testleistungs- bzw. Profildifferenzen:

Die (statistische) Sicherung von Testleistungsdifferenzen folgt dem Meßfehler-Konzept. Über die Bestimmung des Konfidenzintervalls bzw. der Konfidenzintervalle kann jede Maßzahl oder jeder individuelle Testwert hinreichend zuverlässig von anderen Maßzahlen (Testwerten eines anderen Individuums oder früheren Testwerten desselben Individuums) abgegrenzt, d.h. die beobachtete Differenz auf Zufälligkeit versus Überzufälligkeit hin kontrolliert werden. Da jedes Meßergebnis mit Fehlern sowohl des Meßinstrumentes selbst als auch (aus Mängeln) der äußeren und inneren Testsituation behaftet ist, bedarf es zur Sicherung von Testdifferenzen jeweils genauer Angaben über die Reliabilität bzw. den Meßfehler des verwendeten Testinstrumentes. Darüber haben wir uns oben schon eingehend unterhalten (s.S. 63ff.). Die notwendigen Angaben sollten deshalb in jedem Testmanual zu finden sein. Sofern der Meßfehler in Einheiten einer Standardwertskala ausgedrückt ist, kann das gewünschte Vertrauensintervall jeweils leicht im Kopf ausgerechnet werden.

Wesentlich aufwendiger und schwieriger gestaltet sich diese Prozedur, wenn statt *einer* Testskala zugleich mehrere (z.B. Einzeltests einer heterogen zusammengestellten Testbatterie oder Subtests einunddesselben Testsystems) zur Verfügung stehen. Die statistische Sicherung solcher *Testprofile* führt praktisch wieder über die Bestimmung der Konfidenzintervalle der Einzel- oder Subtests. "Für die Deutbarkeit einer Profildifferenz wird gefordert, daß sich die Konfidenzintervalle der beteiligten Tests nicht überschneiden" (Lienert 1967, S. 459). Lienert (loc.cit.) konnte nun zeigen, daß der kritische Unterschied, die sog. *kritische Differenz* ( $d_{crit}$ ), zweier Testprofile wesentlich ökonomischer und zugleich präziser über folgende Beziehung zum Standardmeßfehler von Profildifferenzen ( $s_{ed}$ ) zu errechnen ist:

$$d_{crit} = z_{crit} \cdot s_{ed} \quad \text{beziehungsweise} \\ d_{crit} = z_{crit} \cdot \sigma \sqrt{2 - (r_{11} + r_{22})} \quad , \text{ wobei}$$

$z_{crit}$  = z-Wert ist, der der vorgegebenen bzw. gewünschten Irrtumswahrscheinlichkeit entspricht,

$\sigma$  = Standardabweichung der verwendeten Normenskala (beim LPS bzw. PSB ist  $\sigma = 2$  C) und

$r_{11}$  bzw.  $r_{22}$  die (Retest-)Reliabilitätskoeffizienten der verglichenen (Sub-)Tests symbolisieren.

Die von uns nach dieser Formel<sup>51</sup> berechneten *kritischen Differenzen* beim LPS von Horn

<sup>51</sup> Häufig wird der *Retest*-Reliabilitätskoeffizient in dieser Formel durch einen Konsistenzkoeffizienten ersetzt, was korrekterweise nur dann geschehen sollte, wenn sich Retestreliabilität und Konsistenz nicht wesentlich voneinander unterscheiden. Während die Differenzwerte von Tabelle 3 auf Retestreliabilitäten basieren, legten wir der Berechnung unserer Profildifferenzen in Tabelle 4 die Hornschen Koeffizienten zur Halbierungszuverlässigkeit zugrunde.



finden sich für zwei verschiedene Altersbereiche in den nachstehenden Tabellen 3 und 4 zusammengefaßt, wobei wir uns auf die Reliabilitätskontrolldaten von Tent (1969, S. 87ff.) und Horn (1962, S. 6) stützten. Sofern anstelle der Differenzenmatrix eine "globale" kritische Differenz ( $\bar{d}_{crit}$ ) angegeben werden kann, wird natürlich die Interpretation von Profildifferenzen erheblich vereinfacht. Doch ist dieses Vorgehen nur bei homogenen Einzeltest- bzw. Subtestreliabilitäten erlaubt. Nach unserer  $\chi^2$ -Probe für die LPS-Anwendung in der Grund- und Hauptschule unterscheiden sich jedoch die LPS-Subtestreliabilitäten signifikant voneinander, weshalb hierfür die Erstellung der Differenzenmatrizes unumgänglich wurde. Sonst könnte für  $\bar{d}_{crit}$  = rd. 1 Centilwert angenommen werden.

Tabelle 3

Matrix der Profildifferenzen im LPS (C-Werte) auf der Grundschulstufe (Klasse 4) – Sicherung auf dem 5%-Niveau

LPS	1 + 2	3 + 4	5 + 6	7 - 10	11 + 12	13 + 14	GL
1 + 2	1,4	1,8	1,6	1,6	1,7	1,8	1,5
3 + 4		1,2	1,4	1,4	1,6	1,7	1,3
5 + 6			0,7	1,0	1,2	1,4	0,9
7 - 10				0,7	1,2	1,4	0,9
11 + 12					1,0	1,6	1,1
13 + 14						1,2	1,3
GL							0,5

$$prof^{rtt} = 0,87$$

Tabelle 4

Matrix der Profildifferenzen im LPS (C-Werte in der Anwendung bei Sekundarstufenschülern und Erwachsenen – Sicherung auf dem 5%-Niveau

LPS	1 + 2	3 + 4	5 + 6	7 - 10	11 + 12	13 + 14	GL
1 + 2	0,8	1,5	1,0	0,9	1,1	0,9	0,9
3 + 4		1,2	1,4	1,3	1,5	1,3	1,3
5 + 6			0,6	0,7	1,0	0,7	0,7
7 - 10				0,4	0,9	0,6	0,6
11 + 12					0,8	0,9	0,9
13 + 14						0,4	0,6
GL							0,4

$$prof^{rtt} = 0,7 - 0,9$$

Bevor wir den Gebrauch vorstehender Tabelleninformationen an Hand einiger Anwendungsbeispiele erläutern, sei noch der Begriff der *Profilreliabilität* kurz geklärt. Für die Beurteilung von Testprofilen sind zunächst zwei Voraussetzungen unerlässlich, nämlich einmal ausreichende Reliabilität jedes Einzel- oder Subtests und zum andern niedrige Interkorrelationen der betr. Tests. M.a.W.: Die Profilreliabilität ( $prof^{rtt}$ ) ist umso besser, je reliabler die Einzel- bzw. Subtests und je niedriger ihre Interkorrelationen sind. Bei einer Profilver-

pretation qua individueller (Intelligenz-)Strukturanalyse erhebt sich somit die "Frage, wie groß der Unterschied zwischen zwei Profilpunkten mindestens sein müsse, wenn er nicht mehr als zufallsbedingt angesehen, sondern entsprechend dem Validitätsanspruch der Einzelteste gedeutet werden darf. Selbstverständlich muß man bei relativ geringen Differenzen mit der Möglichkeit rechnen, daß sie zufällig, d.h. durch die Unreliabilität der Einzeltests, entstanden sind und nicht verschiedenen intraindividuellen Merkmalsausprägungen entsprechen" (Lienert 1967, S. 372 f.). Die Profilreliabilität ist demnach unerläßliche Voraussetzung für die Klärung der Frage, ob es sich im gegebenen Falle um ein "echtes" oder nur um ein "Scheinprofil" (wo die Interkorrelationen fast ebenso hoch wie die Einzeltestreliabilitäten sind) handelt. Nach der bei Lienert (loc.cit.) aufgeführten Formel berechneten wir für das LPS folgende Profilreliabilitäten:

$prof_{tt} = 0.93$  bis  $0.91$  resp.  $0.66$  (bei nicht-minderungskorrigierten  $r_{tt}$ -Werten)  
in der LPS-Anwendung bei Sekundarstufenschülern und Erwachsenen;

$prof_{tt} = 0.87$  in der LPS-Anwendung bei Grundschulkindern der 4. Klasse.

Da im allgemeinen Profilreliabilitäts-Koeffizienten von größer als 0.5 als hinreichende Profilzuverlässigkeit gedeutet werden, steht von hier aus einer Profilinterpretation des LPS nichts im Wege. Eine dritte – mehr formale – Voraussetzung für Profil-Testinterpretationen ist beim LPS ohnehin erfüllt: linear transformierte Profilpunkte, d.h. in Standardwerten (C-Werten) ausgedrückte Subtestleistungen.

Mit Hilfe der Differenzenmatrizes zum LPS können nunmehr gesichert auf dem 5%-Niveau (also mit 95%iger Wahrscheinlichkeit versus einem maximalen Fehlerrisiko von 5%) folgende Testinterpretationen vorgenommen werden: 1. Sicherung von *inter*individuellen Differenzen im LPS, 2. Sicherung von *intra*individuellen Differenzen im LPS, 3. Sicherung von *Profil*differenzen im LPS. Einige Beispiele mögen das Vorgehen erläutern.

1. Zwei Viertkläßschüler unterscheiden sich in der LPS-Subtestleistung 5 + 6 tatsächlich, wenn sie mindestens um 0,7 C-Werte auseinanderliegen (vgl. Tab. 3). Für die Interpretation von *inter*individuellen – sowie *intra*individuellen – Profildifferenzen gelten jeweils die Werte in den *Diagonalfeldern* obiger Tabellen. Oder ein anderes Beispiel: Zwei Hauptschüler der 7. Klasse unterscheiden sich im LPS 7-10 bereits dann signifikant, wenn der Testwert des einen Schülers um mindestens 0,4 C höher liegt als der des anderen Schülers (vgl. Tab. 4). Analog kann die Matrix von Tabelle 4 zur Sicherung von Testunterschieden zwischen Haupt- und Realschülern, zwischen Realschülern und Gymnasiasten, zwischen Hauptschülern und Gymnasiasten etc. verwendet werden.

2. Ein Schüler der 4. Grundschulklasse sei zweimal hintereinander, und zwar am Schuljahranfang und am Schuljahrende (die Retestung nach 1 Jahr erfolgte mit der Paralleltestform), mit dem LPS untersucht worden. Zunächst erreichte der Pb eine Gesamtleistung (GL) im LPS von 6,2 C, am Ende desselben Schuljahrs wurden vom gleichen Pb 7,0 C in LPS-GL erzielt. Bei einer notwendigen Mindestdifferenz von 0,5 C (vgl. Tab. 3) ist die zweite Testleistung des Schülers tatsächlich besser als die erste. Hier ein weiteres Beispiel: Ein Gymnasiast habe im Subtest 3 + 4 bei der ersten LPS-Untersuchung 6,8 C und drei Jahre später bei erneuter Testung 7,8 C erzielt. Dieser Unterschied muß als Zufallsergebnis interpretiert werden, da die erforderliche Mindestdifferenz von 1,2 C laut Tabelle 4 hier nicht erreicht wurde.

3. Ein Schüler der 4. Grundschulklasse erreichte 5 C im LPS-Subtest 1 + 2 und 6,4 C im LPS-Subtest 3 + 4. Die Frage, ob der Schüler in seiner Denkleistung besser ist als in seiner verbalen Leistungsfähigkeit, muß hier verneint werden, da der für eine signifikante Differenz notwendige Punkteabstand von mindestens 1,8 C zwischen den Positionen im LPS 1 + 2 und LPS 3 + 4 in unserer Matrix (Tab. 3) nicht erreicht wurde. Andererseits kann z.B. dem Realschüler der 10. Klasse, der 9,0 C im LPS 7-10 und 7,2 C im LPS 5 + 6 erzielte, eine bessere technische Begabung als sprachliche Fähigkeit(en) der im LPS 5 + 6

geforderten Art testiert werden, denn laut Tabelle 4 reichte für eine entsprechende Profil-sicherung bereits eine Differenz von 0,7 C aus.

Die Benutzung der LPS-Differenzenmatrizes ist nicht nur *ökonomischer* als die Berechnung entsprechender Konfidenzintervalle, sie erlaubt auch *präzisere* Schätzungen, was sich indirekt in den vergleichsweise größeren Punktedifferenzen, die nach dem Konfidenzintervall-konzept zur Differenzen-Sicherung notwendig wären, ablesen läßt (vgl. Lienert 1967, S. 463). Die folgenden Tabellen 5 und 6 enthalten nun die entsprechenden Differenzenmatrizes zum PSB von Horn, einer Kurzform des LPS. Die Profilinterpretation des PSB mit Hilfe der Differenzenmatrizes geschieht analog zum LPS (s.S. 158f.). Nach unseren Berechnungen betragen die *Profilreliabilitäten des PSB*

prof<sup>r</sup><sub>tt</sub> = 0.84 (in der Anwendung bei Grundschulern) bzw.

prof<sup>r</sup><sub>tt</sub> = 0.86 (in der Anwendung bei Sekundarstufenschülern und Erwachsenen).

**Tabelle 5** Matrix der Profildifferenzen im PSB (C-Werte) auf der Grundschulstufe (Klasse 4) – Sicherung auf dem 5%-Niveau

PSB	1 + 2	3 + 4	5 + 6	7 + 8	9 + 10	GL	
1 + 2	1,4	1,8	1,6	1,6	1,8	1,5	
3 + 4		1,2	1,5	1,4	1,7	1,3	
5 + 6			0,8	1,1	1,5	1,0	prof <sup>r</sup> <sub>tt</sub> = 0.84
7 + 8				0,7	1,4	0,9	
9 + 10					1,2	1,3	
GL						0,5	

**Tabelle 6** Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei Sekundarstufenschülern bzw. Erwachsenen – Sicherung auf dem 5%-Niveau

PSB	1 + 2	3 + 4	5 + 6	7 + 8	9 + 10	GL	
1 + 2	0,8	1,5	1,0	0,9	0,9	0,9	
3 + 4		1,2	1,4	1,3	1,3	1,3	
5 + 6			0,7	0,8	0,8	0,8	prof <sup>r</sup> <sub>tt</sub> = 0.86
7 + 8				0,4	0,6	0,6	
9 + 10					0,4	0,6	
GL						0,4	

Über weitere Möglichkeiten der Profilanalyse berichtet Lienert (1967, S. 464 ff.). So kann man verschiedene Testprofile noch auf ihre Ähnlichkeit versus ihre Unähnlichkeit hin kontrollieren. Nach denselben Gesichtspunkten kann man Individualprofile mit (diversen) Gruppenprofilen, z.B. Schul- und Studieneignungsprofilen bzw. Schulgruppenprofilen, Berufsgruppenprofilen u.ä. vergleichen. Entsprechende Berechnungen sind aber relativ arbeitsaufwendig, weshalb wir hierauf nicht näher eingehen wollen. Unser Verzicht fällt umso leichter, als anschließend rationellere Verfahren zur (Schul-)Eignungsermittlung ausführlich dargestellt werden. Auch unterließen wir entsprechende Matrix-Berechnungen zum HAWIK bzw. HAWIE, weil die Wechsler-Tests den für eine Profilinterpretation notwendigen statistischen Voraussetzungen nur unzureichend genügen (vgl. unsere Testkritik auf S. 120ff.). Dasselbe gilt für den oben beschriebenen SON.

## Ad b) Probleme und Methoden der Schuleignungsermittlung:

Der Begriff *Schuleignung* meint hier die *überzufällige Wahrscheinlichkeit der Begabung für eine bestimmte Schul- oder Bildungsform*. In Anlehnung an die neuere Begabungsforschung (s.S. 9ff., bes. S. 32ff.) kann dabei *Begabung* als das Insgesamt kognitiver und nichtkognitiver (persönlichkeitspsychologischer) sowie sozio-kultureller Lern- und Leistungsvoraussetzungen definiert werden. Dieser umfassende Begabungsbegriff repräsentiert im wesentlichen jene Faktoren, die Mierke als *Grund- oder Kernintelligenz* sowie als *Hilfs- und Stützfunktionen* (Sekundärintelligenz) eben dieser intellektuellen Grundausstattung des Menschen beschrieben hat. Unser Begriff Schuleignung im Sinne der *Begabung für eine bestimmte schulische Bildungsform*, z.B. die des Gymnasiums, impliziert somit im Hinblick auf die dort geforderten Lernleistungen neben zentralen intellektuellen Voraussetzungen (einem gewissen Ausprägungsgrad oder Niveau der mehr anlagebedingt gedachten Denkfähigkeiten) noch andere Eigenschaften der Persönlichkeit, wie Zuwendungsbereitschaft, Phantasie, Gedächtnis, aber auch Aufmerksamkeit, Konzentration, Belastbarkeit, Arbeitssorgfalt oder Leistungsmotivation, Bildungsinteresse(n) u.ä., die den Anpassungs- oder Stützfunktionen der Intelligenz i.e.S. zugerechnet werden. Diese hinwiederum scheinen in hohem Maße den Umweltbedingungen des Individuums, d.h. schulischen und außerschulischen Komponenten (z.B. Bildungs- und Erziehungseinflüssen) zu unterliegen. Während in der Vergangenheit sehr oft einseitig der biogenetische Anteil der Begabung oder Schuleignung (im oben definierten Sinne) betont und die Umwelt lediglich als Anreizsituation für die "Entfaltung" der Begabungs-"Anlagen" interpretiert wurden, setzt sich heute immer mehr die Erkenntnis durch, daß Begabung und Schuleignung im Sinne bestimmter kognitiver Lern- und Leistungsfähigkeiten wohl annähernd gleichermaßen durch Anlagen und sozio-kulturelle Milieubedingungen bestimmt sein müssen. Die verschiedenen Formen der Begabung und Schuleignung dürfen demnach nicht als starre Mitgift des Individuums verstanden werden; kennzeichnend für solche Phänomene ist vielmehr das *dynamische* Moment der mehr oder weniger erfolgreichen Integration sozial-kultureller Möglichkeiten, im Hinblick auf die Ausbildung verschiedener Schuleignungsformen also die Verzahnung von individuellen Faktoren (Intelligenz- bzw. Persönlichkeitsfaktoren) *und* sozio-kulturellen Determinanten (Erziehungs- und Bildungserfahrungen).

Die Effizienz von Begabung und Schuleignung erweist sich in der Begabungs- (Intelligenz-) bzw. *Schulleistung*. Begabung und Schuleignung sind demnach keine direkt aufweisbaren Größen, ihre Bestimmung ist stets an die Leistungsfunktion gebunden. M.a.W.: Begabung, Intelligenz, Schuleignung u.ä. Faktoren(konstrukte) kognitiver bzw. schulischer Lernfähigkeit (Bildungsfähigkeit) können nur aufgrund manifester Leistungen, z.B. aus Schulleistungen oder (Intelligenz-)Testleistungen, erschlossen werden. Hieraus ergeben sich nun für die psychodiagnostische Praxis in der Schule spezifische Möglichkeiten, denen zugleich eine ganze Reihe methodischer Probleme anhaftet. Die wichtigsten seien im folgenden angesprochen.

Alle Diagnosen versus Prognosen, also auch Intelligenz- oder Begabungsdiagnosen versus Schuleignungsprognosen, basieren auf zwei theoretischen Grundannahmen: der *intraindividuellen Konstanz* menschlicher Verhaltenseigenschaften und der *interindividuellen Differenzen* solcher Verhaltensmerkmale oder Fähigkeiten. Ohne diese Prämissen der innerhalb gewisser Variabilitätsgrenzen auch über längere Zeiträume hinweg gedachten Beständigkeit je individueller kognitiver Leistungsvoraussetzungen und der Tatsache, daß sich die Individuen hinsichtlich des Ausprägungsgrades ihrer Fähigkeiten und Eigenschaften nachweislich voneinander unterscheiden, wären Diagnosen versus Prognosen weder möglich noch sinnvoll. Andererseits betonten wir oben den dynamischen Charakter von Begabung und Schuleignung und damit indirekt die Möglichkeit veränderlicher intellektueller Voraussetzungen in bezug auf die Begabungs- bzw. schulische Lernleistung, was zur postulierten intraindividuellen Merkmalsstabilität als Bedingungskomponente für den diagnostischen Zugang in Widerspruch zu stehen scheint (siehe auch die früheren Erörterungen auf S. 31ff.). Zumindest für längerfristige Prognoseurteile resultieren hieraus unübersehbare Schwierigkeiten. Verschärft stellt sich dieses Problem bei der Schuleignungsermittlung am Ende der

Grundschulzeit, wo über einen Zeitraum von 5 oder 6 bzw. 9 Jahre hinweg gültige Begabungsprognosen erforderlich werden, d.h. unter Berücksichtigung wahrscheinlicher – persönlich-motivationaler *und* sozio-kultureller – Veränderungseffekte (künftige) Schulerfolge vorausgesagt werden sollen. Was sich somit als methodisches Problem treffsicherer Schuleignungsprognosen darstellt, erweist sich freilich von nicht geringer pädagogischer Bedeutung. In der modernen Bildungsforschung spricht man in diesem Zusammenhang direkt von der Möglichkeit des "Begabens", also der sozio-kulturellen Einflußmöglichkeit auf die Entwicklung von Begabung und Schuleignung. Die Vermittlung sprachlicher Anregungsmuster in der Frühkindheit, der Aufbau von Lern- und Leistungsmotivationen, die Weckung von Bildungsinteressen oder die Erziehung zu einer positiven Arbeitshaltung sind solche Modi einer optimalen Begabungsentfaltung bzw. Begabungsvergrößerung, wie wir sie bereits aufgewiesen haben (s.S. 35ff.).

Bei der *diagnostischen Erfassung der Schuleignung(en)* rücken prinzipiell zwei Aspekte in den Blickpunkt: einmal je bestimmte intellektuelle Fähigkeiten und persönlichkeitspezifische Eigenschaften des Schülers, zum andern von den Bildungszielen und Aufgaben der in Frage kommenden Bildungseinrichtungen (im dreigliedrigen Schulsystem: Hauptschule, Realschule, Gymnasium) je konkrete Leistungsanforderungen, auf die bezogen bestimmte Eigenschafts- bzw. Fähigkeitsstrukturen als individuelle Voraussetzungen schulischen Bildungserfolges empirisch aufweisbar sein müssen. Hinzu kommt schließlich die Berücksichtigung außerschulischer, d.h. in erster Linie familialer resp. bildungssozialer, Komponenten des Begabens und damit der Konstituierung von Schuleignungen. Da jedoch Termini wie Begabung und Schuleignung letztlich nur theoretische Konstrukte eines wie auch immer sich manifestierenden Leistungsverhaltens, z.B. in der Schule und im Beruf, repräsentieren, bedeutet dies für ihren operationalen Zugriff, daß stets vom aktuellen (zum Zeitpunkt der Untersuchung *gegenwärtigen*) Leistungsverhalten des Schülers (in der Grundschule, im Intelligenztest etc.) auf sein potentielles (*zukünftiges*) Leistungsverhalten geschlossen werden muß. Somit läßt sich jetzt präzisierend der Begriff der *Schuleignung* (z.B. für weiterführende schulische Bildungseinrichtungen) im Sinne von *operational bestimmbaren Erfolgs- und Bewährungswahrscheinlichkeiten* (für Gymnasium, Realschule, Hauptschule) definieren (vgl. Heller 1970 a, 1970 b).

Analog dazu könnte man den Begriff der *Sonderschuleignung* formaliter festlegen als "Eignung" für diese oder jene Sonderschulform. Allerdings müßte zu diesem Zweck der Eignungsbegriff über seinen mehr intellektuell gefärbten Bedeutungskern hinaus erweitert werden, zumal hier vielfach noch der Heilungsaspekt (vgl. *Heilpädagogik*) oder – wertneutraler ausgedrückt – die mehr oder weniger zwingende Notwendigkeit sonderpädagogischer, d.h. *spezieller* Erziehungs- und Bildungsmaßnahmen hervortritt. Der Begriff "*Sonderschulbedürftigkeit*" ist m.E. freilich auch keine glückliche Formulierung. Intelligenzdiagnostische Probleme im Rahmen sonderpädagogischer Fragestellungen werden im nächsten Abschnitt erörtert.

Treffsichere Schulleistungsvoraussagen im Hinblick auf eine bestimmte Bildungsform, d.h. verlässliche und valide *Schuleignungsprognosen*, sind ohne hinreichend gesicherte Beurteilungsmaßstäbe, also die Kenntnis einschlägiger *Schuleignungskriterien*, undenkbar. Die Gewinnung solcher Eignungskriterien bereitet allerdings nicht geringe Schwierigkeiten. Aus zahlreichen Untersuchungen zur diagnostischen Valenz der üblicherweise verwendeten Prädiktorvariablen (z.B. Lehrerurteile in Form von Zeugnisnoten, Grundschulgutachten, sog. Probearbeiten in der 4. Grundschulklasse oder Aufnahmeprüfungen für den Eintritt in die Realschule bzw. das Gymnasium) wissen wir, daß diese vielfach den erforderlichen Meßgütekriterien (Objektivität, Reliabilität, Validität) nicht oder höchst unzulänglich genügen. Am ehesten scheint eine Kombination von Lehrer- und Testurteilen (Grundschulzeugnisse und Tests zur Erfassung der Intelligenz einschließlich relevanter Hilfs- und Stützfunktionen) treffsichere Schuleignungsprognosen zu einem gerade noch vertretbaren Maß zu ermöglichen, wenngleich bei (optimalen) Übereinstimmungskoeffizienten von Eignungsvorhersage und Schulerfolg um 0.5 und 0.6 kaum die Hälfte der Varianz aufgeklärt ist.

Eine weitere Schwierigkeit für valide und reliable Schuleignungsprognosen liegt in der Tatsache *uneinheitlicher Anforderungsmaßstäbe* selbst innerhalb identischer Schulbildungsformen (z.B. intragymnasial), die wiederum zur Unsicherheit schulischer Leistungsbewertung (Zensurierung) qua Bezugskriterium des Schulerfolgs in Relation stehen.

Bei der Erörterung der *Kriteriumskonsistenz* – damit ist die Zuverlässigkeit sog. Außenkriterien (Lehrerurteile, Schulzensuren u.ä.), hier des Schul- und Bildungserfolgs auf Gymnasium, Real- und Hauptschule, angesprochen – spielt noch die Frage nach der *Auswirkung möglicher Bildungsreformen* eine wichtige Rolle. Wenn Begabung und Schuleignung immer auch als vom sozio-kulturellen Milieu abhängige *Variablen* gedeutet werden müssen, dann können grundlegende Veränderungen unseres derzeitigen Schulsystems, der vorschulischen und familiären Erziehungssituation nicht ohne Wirkung auf die Entwicklung von Begabung und Schuleignung bleiben. Ungenaue Zielvorstellungen solcher Reformbestrebungen resp. die Unkenntnis in die Zukunft reichender Aktivitäten erhöhen natürlich die Schwierigkeiten einer treffsicheren Schuleignungsprognose beträchtlich, zumal sich deren Richtigkeit versus Fehlschätzung erst in späteren Jahren erweist und dann kaum noch Korrekturen möglich sind.

Die Vorhersagbarkeit individueller Schulerfolge ist jedoch nicht nur von “äußeren” Faktoren (sozio-kulturellen Determinanten) abhängig, die Varianz individuellen Leistungsvermögens erscheint nicht weniger abhängig von der persönlichen Motivations- und Interessenlage, von den intellektuellen Voraussetzungen im engeren Sinne ganz abgesehen. Die Erscheinungsformen des *Underachievement* (d.h. in bezug auf die intellektuellen Lernleistungsvoraussetzungen deutlich *unter* der Erwartung bleibende Schulleistungen) und des *Overachievement* (d.h. in bezug auf die intellektuellen Lernleistungsvoraussetzungen deutlich *über* der Erwartung liegende Schulleistungen) sind berechte Beispiele für eine solche These. Freilich muß man auch diese Phänomene in gewisser Interdependenz zur sozial-kulturellen Situation in Schule und Elternhaus sehen (vgl. ausführlicher Heller 1969 a, 1969 b, 1970 a und Kemmler 1967, bes. S. 97ff.).

Sieht man von extremen Verschiebungen im äußeren und inneren Kräftespiel kognitiver Leistungsvollzüge ab, dann wird man im allgemeinen davon ausgehen können, daß mit zunehmendem Alter ansteigend eine gewisse Stabilität individueller Begabungsleistungen hinsichtlich Niveau und Differenzierungsgrad erreicht wird. Spätestens am Ende der Grundschule sind demnach – prinzipiell – Schuleignungsprognosen möglich. Ob sich die Sicherheit entsprechender Prognosen zu einem späteren Zeitpunkt, etwa am Ende der Förderstufe oder der Sekundarstufe I, steigern läßt, bleibt abzuwarten. Vielfach gehegte Hoffnungen in dieser Richtung müssen in jedem Falle empirisch ver- oder falsifiziert werden; andererseits sollte die Entscheidung für oder gegen ein bestimmtes Schulsystem nicht zuallererst von (wie auch immer aussehenden) Selektionskriterien abhängig gemacht werden, hier prävalieren eindeutig pädagogische Gesichtspunkte. Daraus darf aber nicht abgeleitet werden, daß in Zukunft Intelligenztestungen oder Schuleignungsermittlungen überflüssig werden, sondern nur, daß sich diese pädagogischen Prinzipien unterzuordnen haben. Zur Erfüllung des pädagogischen Auftrags werden mehr denn je objektive, zuverlässige und gültige Aussagen über Schülerbegabungen und Eignungen für differenzierte Bildungsangebote notwendig sein.

Unsere bisherigen Methodenerörterungen rückten zwei Gesichtspunkte in den Vordergrund der Betrachtung: das Problem der intraindividuellen *Merkmalskonstanz* und die sog. *Kriteriumskonsistenz*, d.h. die Forderung nach systemeinheitlichen (objektiven) Leistungs- und Bewertungsmaßstäben der Schule bzw. Bildungseinrichtungen. Schließlich ergeben sich noch Probleme der *Prädiktorvariablen* selbst, die folgenden Anforderungskriterien entsprechen müssen: a) der *Objektivität* oder Vergleichbarkeit und somit Kontrollierbarkeit zugrundegelegter Maßstäbe bei der Schuleignungsfestlegung; b) der *Validität*, womit letztlich die Richtigkeit der Vorhersage angesprochen ist; c) der *Repräsentanz* schulischer Anforderungskriterien in den gewählten Prädiktorvariablen bzw. Meßinstrumenten (z.B. Intelligenztests), d.h. die zur Schuleignungsermittlung verwendeten Verfahren (Tests,

Fragebogen usw.) müssen den gesamten Bereich der Grund- oder Kernintelligenz sowie die für den Schulerfolg relevanten Stütz- und Hilfsfunktionen der Intelligenz einschließlich des sozio-kulturellen Kontextes in seinen wichtigsten Bedingungsvariablen abdecken; d) schließlich müssen Prädiktorvariablen ausreichende *Reliabilität* (Stabilität bzw. Konsistenz) besitzen. Von Prädiktoren, die diesen Gütekriterien einer Messung nicht oder nur mangelhaft genügen, kann nach allen vorliegenden Erfahrungen keine treffsichere Schuleignungsprognose erwartet werden. Dieses Urteil trifft insbesondere Lehrerurteile der herkömmlichen Form, also Grundschulgutachten, Zensuren, nichtstandardisierte Probearbeiten, Aufnahmeprüfungen u.ä. Somit kommen als brauchbare Instrumente resp. Prädiktoren zur Begabungsdifferenzierung resp. Schuleignungsprognose praktisch nur vollstandardisierte Intelligenz- und Leistungstests infrage, die im Hinblick auf ein gleich noch zu erörterndes statistisches Problem möglichst faktorisierte Verfahren darstellen sollten. Zum Problemkomplex der sog. Übertrittsauslese haben wir andernorts (Heller 1969 b, S. 390 ff. und 1970 a, S. 43 ff.) ausführlicher Stellung genommen; siehe u.a. noch die umfassenden Darstellungen bei Schultze (1964), Tent (1969) und Undeutsch (1969). Die folgende Diskussion wird sich mehr auf die neueren Verfahrensansätze der Begabungsdifferenzierung bzw. Schuleignungsermittlung konzentrieren, wobei die Gewinnung reliabler und valider Schuleignungskriterien im Mittelpunkt des Interesses stehen soll.

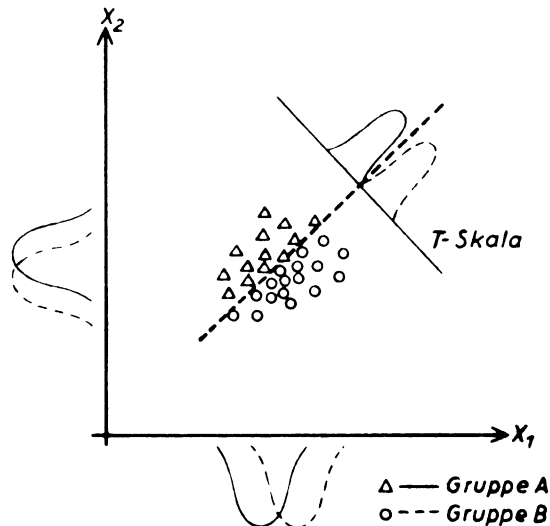
Das Methodenproblem Nr. 1 einer treffsicheren Schuleignungsprognose besteht darin, ein Klassifikationsverfahren zu finden, das eine optimale Trennung heterogener Begabungsgruppen erlaubt. Nach wie vor am häufigsten kommt hierbei die sog. *Grenzwertmethode* zur Anwendung: Innerhalb einer empirisch (z.B. testdiagnostisch) ermittelten Intelligenzverteilung wird mehr oder weniger willkürlich ein Grenzwert festgelegt; die mit ihren Testleistungswerten über versus unter diesem *Cutting Score* gelegenen Probanden repräsentieren dann zwei verschiedene Begabungs- oder Schuleignungsgruppen und markieren in ihren Verteilungsgrenzen zugleich die entsprechenden Eignungskennbereiche. Diesem Verfahren haften jedoch erhebliche Nachteile an. So ergeben sich fallweise sehr unterschiedliche Eignungsquoten je nach der Höhe des festgesetzten Cutting Score. Bei einem Grenzwert von IQ = 115 wird beispielsweise die Gymnasialeignungsquote wesentlich kleiner ausfallen als bei einer unteren Grenze von IQ = 110. Zudem streut die Meinung der Experten bezüglich der Höhe des kritischen Eignungswertes erheblich. Ferner muß darauf hingewiesen werden, daß ein einzelner Kennwert (z.B. Gesamt-IQ) oder auch mehrere Testdurchschnittswerte dem variablen Bedingungsgeflecht bestimmter Schuleignungsformen nur unzulänglich gerecht werden. Damit in Zusammenhang steht der gravierendste Nachteil aller Grenzwertmethoden: die starke Überlappung einzelner Begabungs- und Schuleignungsgruppen (siehe unten). Der beschriebene Ansatz stellt somit allenfalls ein sehr grobes und in vielerlei Hinsicht recht unzulängliches Klassifikationsmodell dar.

Neben der Grenzwertmethode und ihren zahlreichen Varianten verdient ein erst in jüngster Zeit unter der Bezeichnung "Approximative Expertenabstimmung" bekannt gewordener Verfahrensansatz im Zusammenhang unserer Problemerkörterung Bedeutung. Als Spezifikum der *Approximativen Expertenabstimmung* kann nach Aurin (1968, S. 30 f.) eine spezifische Beratungsinteraktion von Lehrer(n) und Psychologe(n) bzw. Bildungsberater(n) benannt werden. Die auf diese Weise erarbeiteten Schuleignungsurteile resp. Bildungsempfehlungen vereinigen sowohl Test- als auch Schul- bzw. Lehrerurteilkriterien mit wechselnder, meist stärkerer Gewichtung der Testprädiktoren. Das Verfahren hat sich in der baden-württembergischen Bildungsberatung im großen und ganzen bewährt. Es erfordert allerdings ein gut eingearbeitetes Beraterteam und ist zudem bei größeren Begabungsuntersuchungen verhältnismäßig zeit- und arbeitsaufwendig, also unökonomisch. Auch machen sich gelegentlich unerwünschte subjektive Urteileinflüsse bemerkbar, ohne daß diese allerdings bisher zum Kardinalproblem wurden. Erster wiegt hier der Einwand, daß via *Approximative Expertenabstimmung* Begabungs- und Schuleignungsuntersuchungen ab einer bestimmten Größenordnung, etwa bei Jahrgangspopulationen, praktisch (wegen des erforderlichen personellen und finanziellen Aufwandes) nicht mehr realisierbar seien.

Hieran knüpften nun unsere eigenen Überlegungen zur rationellen und zugleich objektiven, zuverlässigen und validen Erfassung von Begabungen resp. Schuleignungen<sup>52</sup>.

Der zunächst geplante Ansatz, über *Regressionsanalysen* die erforderlichen Schuleignungskriterien für die Begabungsdifferenzierung zu gewinnen, mußte schon bald in Anbetracht der unzuverlässigen Bezugskriterien (z.B. aufgrund unterschiedlicher Notenmaßstäbe irreliablen Schulzensuren) wieder aufgegeben werden. Auf der Suche nach einem von (unzuverlässigen) Außenkriterien unabhängigen Wege der Schuleignungsermittlung wurde schließlich ein internes Kriterien folgendes multivariates Klassifikationsverfahren in Betracht gezogen. Erste Versuche, die mit Hilfe der *multiplen Diskriminanzanalyse* an Hand eines kleineren Schülersamples bereits 1965 durchgeführt wurden, bestätigten die Erwartungen bezüglich einer optimalen Schulgruppendifferenzierung recht ermutigend (vgl. Aurin 1966, S. 147 ff.). Zur theoretischen und methodologischen Kennzeichnung des Verfahrens verweisen wir auf die einschlägige Statistikk-literatur, z.B. Cooley & Lohnes (1962), Weber (1967) oder Überla (1968); siehe neuerdings auch Allinger & Heller (1972). Die Funktion der multiplen Diskriminanzanalyse erweist sich in der "Trennung verschiedener Gesamtheiten und Zuordnung fraglicher Elemente zu einer der Gesamtheiten" (Weber). Im weiteren Verlauf erprobten wir schließlich in Zusammenarbeit mit dem DRZ (Deutschen Rechenzentrum) in Darmstadt das IBM-Programm AUKL (*Automatische Klassifikation*); diese Klassifikationstechnik basiert auf der multiplen Diskriminanzanalyse. Die Vorzüge der AUKL gegenüber den oben beschriebenen Verfahren zur Schuleignungsermittlung sind mannigfacher Art.

Eines der Hauptprobleme bei Verwendung der Grenzwertmethode oder einer ihr verwandten Methode resultiert aus dem Phänomen des Overlapping von Testleistungswerten unterschiedlicher Provenienz, hier verschiedener Schulleistungs- bzw. Schuleignungsgruppen. Mit Hilfe der Diskriminanzanalyse oder der Automatischen Klassifikation ist es nun möglich, solche Überschneidungen einzelner Schul- bzw. Testleistungsbereiche auf ein Minimum zu reduzieren und damit eine optimale Gruppentrennung (Klassifikation verschiedener Schuleignungsgruppen) zu gewährleisten. Dieser Effekt beruht praktisch auf einer Maximisierung der Varianz zwischen den zu klassifizierenden Gruppen (in unserem Falle zwischen Gymnasial-, Real- und Hauptschuleignungen) und einer Minimalisierung der Varianz innerhalb dieser Gruppen. Bei der Klassifikation in drei oder mehr Gruppen sind oft mehrere Trennfunktionen zum Erhalt optimaler, d.h. homogener Gruppenverhältnisse notwendig. Aus der abgebildeten Veranschaulichungsskizze (entnommen K. Überla 1968, S. 87) zur Diskriminanzanalyse wird deutlich, wie die Projektionen der Testwerte zweier Untersuchungsgruppen auf die Achsen  $X_1$  und  $X_2$  zwei sich überlappende Verteilungen ergeben, während in der Projektion auf die T-Skala beide Personengruppen getrennt erscheinen.



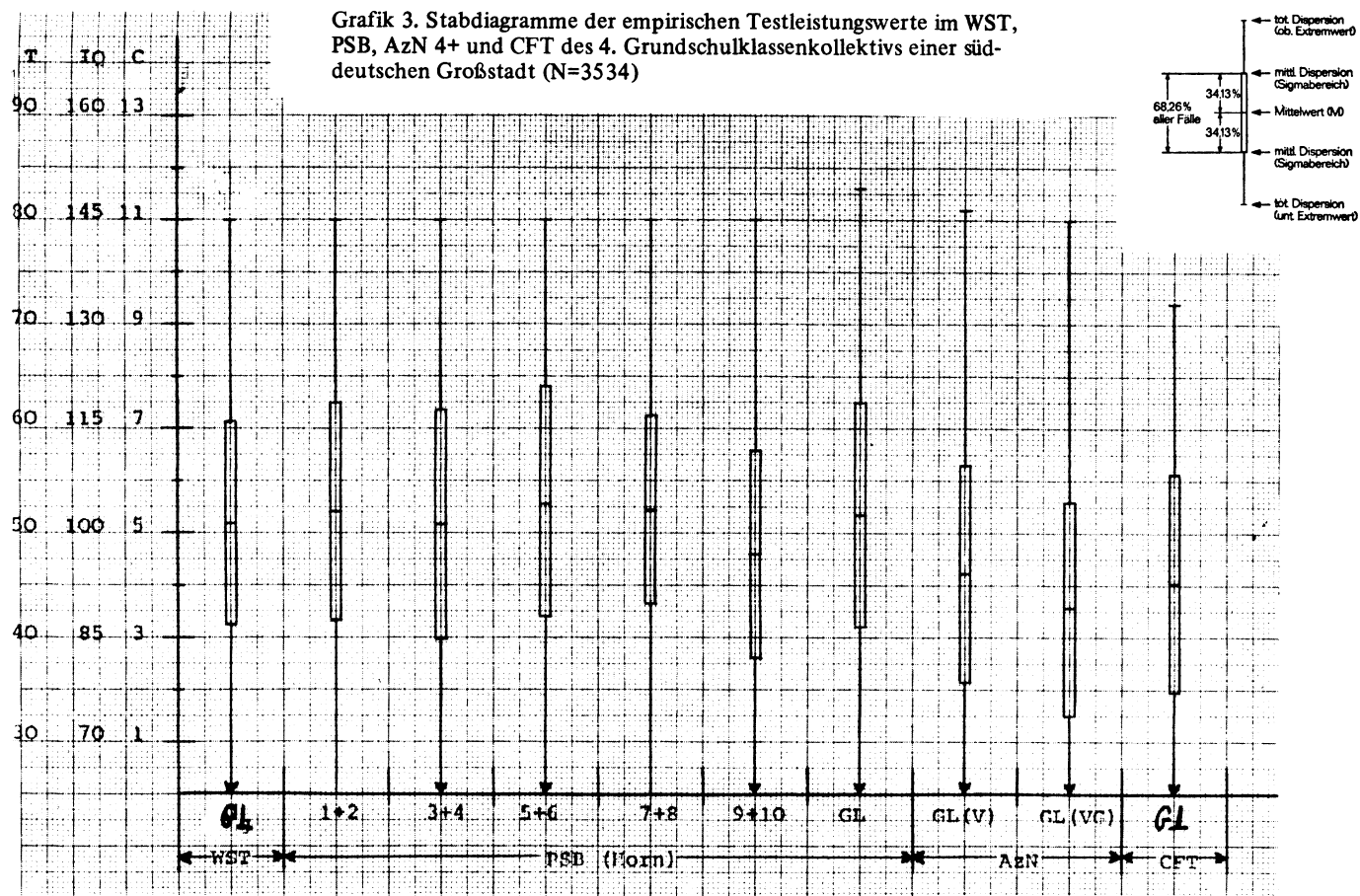
<sup>52</sup> Nachstehende Erörterungen erfolgen in enger Anlehnung an Heller (1970 a, S. 106 ff.).

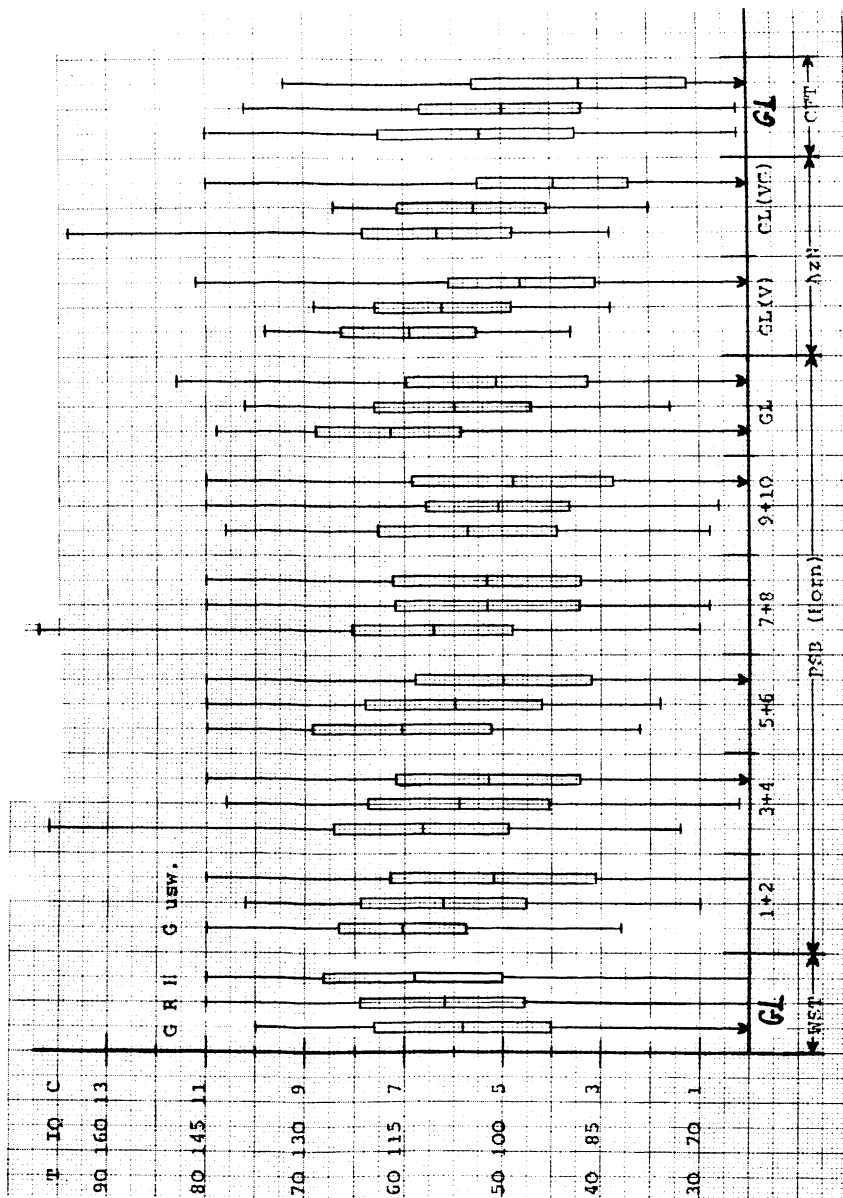


Tabelle 7  
Mittelwerte, Sigtabereiche und Variationsbreite (Extremwerte) in C der nach AUKL ermittelten Schuleignungsgruppen für  
Hauptschule, Realschule und Gymnasium im LPS

Sample	N	1 + 2	3 + 4	5 + 6	7 - 10	L - P - S 11 + 12	13 + 14	13	GL	APR
Kl. 6 H	840	3,7 2,2 - 5,2 (0,0 - 8,0)	3,8 2,2 - 5,4 (0,0 - 9,0)	3,6 2,1 - 5,1 (-1,0 - 9,0)	5,0 3,1 - 6,9 (0,0 - 10,0)	4,2 2,4 - 6,0 (0,0 - 10,0)	4,0 2,1 - 5,9 (-0,2 - 12,0)	4,3 2,4 - 6,2 (-1,0 - 9,0)	4,4 2,8 - 6,0 (0,0 - 8,6)	4,6 2,9 - 6,3 (-1,0 - 12,0)
Kl. 8 H	657	4,2 2,9 - 5,5 (-1,0 - 8,0)	4,3 2,6 - 6,0 (-0,2 - 10,0)	4,0 2,6 - 5,4 (-0,2 - 8,0)	5,2 3,3 - 7,1 (-1,0 - 10,0)	4,8 3,2 - 6,4 (-1,0 - 10,0)	4,5 2,5 - 6,5 (-1,0 - 11,0)	4,1 2,1 - 6,1 (0,0 - 9,0)	5,0 3,6 - 6,4 (0,0 - 9,8)	4,7 2,9 - 6,5 (-1,0 - 11,0)
Kl. 6 R	42	6,1 5,1 - 7,1 (4,0 - 9,0)	5,5 3,7 - 7,3 (3,0 - 10,0)	6,1 4,7 - 7,5 (3,0 - 9,0)	6,9 5,2 - 8,6 (3,0 - 11,0)	6,8 5,0 - 8,6 (1,0 - 10,0)	5,6 3,8 - 7,4 (2,0 - 9,0)	4,4 2,5 - 6,3 (1,0 - 8,0)	7,1 5,5 - 8,7 (3,8 - 12,0)	6,2 4,3 - 8,1 (3,0 - 11,0)
Kl. 8 R	40	6,2 5,3 - 7,1 (5,0 - 8,0)	5,9 3,9 - 7,9 (1,0 - 10,0)	5,8 4,1 - 7,5 (2,0 - 9,0)	6,1 4,4 - 7,8 (2,0 - 9,0)	6,3 4,8 - 7,8 (3,0 - 9,0)	6,1 4,0 - 8,2 (2,0 - 10,0)	4,7 2,4 - 7,0 (0,2 - 9,0)	6,8 5,4 - 8,2 (3,8 - 10,2)	6,2 4,4 - 8,0 (3,0 - 11,0)
Kl. 10 R	130	6,7 5,7 - 7,7 (4,0 - 9,0)	6,4 4,6 - 8,2 (3,2 - 11,0)	6,1 4,5 - 7,7 (3,0 - 9,0)	6,7 4,9 - 8,5 (2,9 - 10,0)	6,5 4,8 - 8,2 (3,0 - 10,0)	5,6 3,6 - 7,6 (2,0 - 9,8)	4,6 2,5 - 6,7 (1,0 - 9,0)	7,1 5,8 - 8,4 (4,0 - 11,0)	5,8 4,0 - 7,6 (2,8 - 11,0)
Kl. 6 G	106	6,7 5,2 - 8,2 (4,0 - 11,0)	7,0 5,4 - 8,6 (5,0 - 11,0)	7,3 5,8 - 8,8 (3,0 - 11,0)	7,6 6,1 - 9,1 (3,0 - 12,0)	7,0 5,3 - 8,7 (3,0 - 11,0)	6,5 4,4 - 8,6 (2,0 - 13,0)	4,0 2,1 - 5,9 (0,0 - 8,0)	8,3 7,1 - 9,5 (4,6 - 11,4)	6,7 5,2 - 8,2 (3,0 - 12,0)
Kl. 8 G	112	7,6 6,6 - 8,6 (5,0 - 10,0)	7,0 5,3 - 8,7 (3,0 - 12,0)	7,0 5,5 - 8,5 (4,0 - 10,0)	7,1 5,5 - 8,7 (3,0 - 11,0)	6,9 5,2 - 8,6 (3,0 - 11,0)	5,5 3,9 - 7,1 (3,0 - 9,0)	4,3 2,1 - 6,5 (0,0 - 10,0)	8,1 6,5 - 9,7 (3,0 - 11,8)	5,6 4,1 - 7,1 (3,0 - 10,0)
Kl. 10 G	111	7,8 6,8 - 8,8 (5,0 - 10,0)	7,5 5,9 - 9,1 (3,0 - 11,0)	7,4 6,0 - 8,8 (4,0 - 10,0)	6,7 5,0 - 8,4 (3,0 - 10,0)	6,9 5,4 - 8,4 (3,0 - 10,0)	5,6 3,6 - 7,6 (1,0 - 11,0)	4,1 1,7 - 6,5 (0,0 - 9,0)	8,1 6,6 - 9,6 (4,2 - 11,4)	5,8 3,8 - 7,8 (2,0 - 14,0)
Kl. 12 G	146	7,5 6,6 - 8,4 (5,0 - 10,0)	7,3 5,6 - 9,0 (4,0 - 10,0)	6,8 5,6 - 8,0 (5,0 - 10,0)	7,1 5,6 - 8,6 (3,0 - 10,0)	6,3 4,9 - 7,7 (3,0 - 10,0)	5,6 4,1 - 7,1 (2,0 - 10,0)	4,7 2,6 - 6,8 (0,0 - 9,0)	7,9 6,5 - 9,3 (5,2 - 11,4)	5,8 4,3 - 7,3 (3,0 - 11,0)

Grafik 3. Stabdiagramme der empirischen Testleistungswerte im WST, PSB, AzN 4+ und CFT des 4. Grundschulklassenkollektivs einer süd-deutschen Großstadt (N=3534)





Grafik 4. Stabdiagramme der empirischen Testleistungswerte im WST, PSB, AzN 4+ und CFT dreier Schultichproben des 5. Schuljahrs einer süddeutschen Großstadt (WST- und AzN-Werte resultieren aus Schultyp-(Klassen-)Normen, PSB- und CFT-Werte resultieren aus Altersnormen) Legende: G = Gymnasiasten (N=110); R = Real-schüler (N=105); H = Hauptschüler (N=119 bzw. 92)

Allgemein ist die Diskriminierung verschiedener Leistungs- oder Eignungsgruppen umso effektiver, je besser die einzelnen Testleistungsvariablen bereits *vor* der automatischen Klassifikation (AUKL) die betreffenden Gruppen trennen und je niedriger die Interkorrelationen der verwandten Tests ausfallen. Auch erweisen sich *faktorierte* Tests, d.h. im Hinblick auf ein bestimmtes Faktorenmodell konstruierte (und kontrollierte) Tests, nichtfaktorierten in der Regel überlegen in bezug auf AUKL (vgl. noch Janke 1964). Weitere Einzelheiten über das AUKL-Verfahren finden sich in den Publikationen von Heller (1970 a) und Allinger & Heller (1972) sowie der dort angegebenen Literatur, besonders von Faber & Nollau (1969).

Die Überlegenheit des AUKL-Verfahrens gegenüber den herkömmlichen Methoden der Begabungs- und Schuleignungsdifferenzierung zeigt sich aber nicht nur in einer den jeweiligen Eignungsverhältnissen besser angepaßten Trennfunktion, der AUKL-Ansatz bietet auch die Möglichkeit, *objektive* resp. (z.B. für alle Schüler einer bestimmten Klassenstufe eines Bundeslandes) *einheitliche Beurteilungsmaßstäbe* hinsichtlich der Schuleignungsfindung oder Begabungs-/Leistungsdifferenzierung (etwa innerhalb der Gesamtschule) zu verwirklichen. Darüber hinaus bietet AUKL als rationelles Klassifikationsverfahren wesentliche ökonomische und zeitliche Vorzüge wie kein anderes derzeitig praktiziertes Differenzierungsverfahren. Was schließlich die *Validität*, also die Vorhersagegültigkeit des AUKL-Verfahrens betrifft, so belegen die Ergebnisse umfangreicher Bewährungs- und vergleichender Methodenkontrollen (Heller 1970 a, 1972 c; Allinger & Heller 1972) bei unkorrigierten Validitäts- bzw. Übereinstimmungskoeffizienten zwischen  $r = 0.5$  und  $0.7$  sowie korrigierten Werten zwischen  $r = 0.6$  und  $0.8$  praktisch – sub specie Methode – optimale Prognosegültigkeit, wie sie bislang nur selten erzielt wurde. Wir stimmen deshalb Tent (1969, S. 131 ff.) zu, wenn er sich weitere methodische Verbesserungen nur über den Weg von Systemverbesserungen erhofft; diese wären z.B. aufgrund objektiverer Schulleistungsbeurteilungen, einheitlich definierter Lernziele im schulischen Bereich u.dgl.m. denkbar.

An Hand der via AUKL eingeteilten "reinen" Schuleignungsgruppen – im folgenden kurz "AUKL-Gruppen" genannt – lassen sich nun jeweils zugehörige Testleistungswerte bestimmen, die als operational gewonnene *Schuleignungskriterien* definiert werden können. Die auf repräsentative Schülersamples bezogenen Testwerte in Form von Leistungskennbereichen im LPS, WST (FWT), MTVT und BIT finden sich, nach Schultyp und Klassenstufe (6-12) getrennt aufgeführt, in unserer schon mehrfach erwähnten Buchpublikation (Heller 1970 a, S. 127 ff., vgl. bes. Graf. 10-20) dargestellt. Hier sind die wichtigsten Ergebnisse in Tab. 7 zusammengefaßt wiedergegeben.

Inzwischen konnten wir entsprechende Eignungskennbereiche für weitere Testverfahren (PSB, AzN 4+, CFT und wiederum WST) sowie die – empirischen – Schulgruppen der Klasse 4 (Grundschule) und 5 (Hauptschule, Realschule, Gymnasium) bestimmen. Die in den Grafiken 3 und 4 dargestellten Parameter (Mittelwerte und Dispersionsmatrizen der *empirischen* – nicht der AUKL-differenzierten – Schulgruppen) beziehen sich diesmal ausschließlich auf *Großstadtpopulationen* und sind vor allem als Hilfen für die Begabungsdiagnostik in der Schule gedacht. Die Verteilungsparameter der einzelnen Schulgruppen deuten an, innerhalb welcher Testleistungsbereiche (Eignungskennbereiche) sich schulische Begabungen bewähren und entwickeln können.

In den (hier nicht wiedergegebenen) Grafiken der *BIT-Profile* unserer via AUKL differenzierten Gymnasial-, Real- und Hauptschuleignungsgruppen kommen zwar gewisse Interessenunterschiede zwischen den Schulgruppen, besonders in den jeweiligen Abschlußklassen, zum Ausdruck, doch muß insgesamt die berufsorientierende Funktion des BIT höher veranschlagt werden als seine schuleignungsdiakritische (vgl. Heller 1970 a, S. 157 ff. u. S. 193 ff.). Immerhin lassen sich für die drei Eignungsgruppen nachstehende Rangfolgen bevorzugter (Berufs-)Interessen im BIT angeben:

**Gymnasiasten** bzw. G-Eignungsgruppe: LG, GH, EH, SE, TN (bzw. TN, SE), sowie LF, KB, VB und TH (am wenigsten gefragt);  
**Realschüler** bzw. R-Eignungsgruppe: GH, EH, LG, KB, LF, TN (bzw. TN, LF), sowie SE, VB und TH (am wenigsten gefragt);  
**Hauptschüler** bzw. H-Eignungsgruppe: GH, EH, KB, SE, LG, VB, LF, TH, TN.

Sieht man einmal von den wenig diakritischen BIT-Dimensionen GH und EH ab, dann wird deutlich, daß die *Gymnasiasten* vor allem ausgeprägte literarisch-geistige, sozial-erzieherische und technisch-naturwissenschaftliche Interessen haben, während kaufmännische, Verwaltungs- und handwerkliche Tätigkeiten hier kaum gefragt sind. Die *Hauptschüler* hingegen verfügen über stärkere Interessen für kaufmännische, sozial-erzieherische, literarisch-geistige und Verwaltungsberufe, wohingegen technisch-naturwissenschaftliche Betätigungen hier das geringste Interesse finden. Die *Realschüler* schließlich nehmen mehr eine Mittelstellung zwischen diesen beiden Schülergruppen ein, wenngleich sie hinsichtlich ihrer Interessenprofilierung den Hauptschülern näher stehen als den Gymnasiasten, die überhaupt die markantesten BIT-Profile aufweisen.

Bezüglich der *Intelligenztestprofile* (vgl. M-Werte in Graf. 4) ergeben sich allerdings eher umgekehrte Verhältnisse, d.h. hinsichtlich Grad und Ausprägung intellektueller Fähigkeiten unterscheiden sich die Realschüler stärker von den Hauptschülern als von den Gymnasiasten – ausgenommen die PSB-Dimensionen 7 + 8 sowie 9 + 10. Die im WST "umgedrehten" Profiltendenzen widersprechen den übrigen Testleistungstrends der drei Schuleignungsgruppen nur scheinbar; sie erklären sich als methodischer Artefakt inadäquater Schultypnormen. In der unausgewählten Grundschulpopulation (vgl. Graf 3) tritt dieser Effekt nicht auf. Allerdings machen sich dort teilweise (etwa im lernleistungs-relevanten AzN 4+) die zum Zeitpunkt der Datenerhebung (1968) laufenden Kurzschuljahre bemerkbar, worauf wir in einer in Kürze erscheinenden Veröffentlichung näher eingehen werden (vgl. Heller 1973).

Den Stabdiagrammskizzen zum PSB, AzN und CFT (Graf. 4) sowie den entsprechenden Profilwerten zum LPS (Tab. 7) ist zu entnehmen, daß sich G-, R- und H-Eignungen in den genannten Testleistungsdimensionen mehr oder weniger deutlich voneinander abheben. Zur Unterscheidung tragen vor allem folgende – dem Gewicht ihrer Diskriminanzfunktion nach geordnete – Testprädiktoren bei (vgl. Allinger & Heller 1972, S. 18 ff.):

LPS 1 + 2, GL, 5 + 6, 3 + 4, 11 + 12, 7 – 10; PSB 1 + 2, GL, 5 + 6, 3 + 4, 9 + 10, 7 + 8; ferner AzN 4+ und (in geringerem Maße) CFT.

Die interschulischen Testleistungsdifferenzen (besonders der AUKL-Gruppen) in bezug auf die genannten Prädiktorvariablen sind fast durchweg signifikant, nicht dagegen die (gelegentlich) beobachteten intraschulischen Differenzen (Heller 1970 a, S. 149 ff.). Daraus leiten wir folgende Kennzeichnungen für die Gymnasial-, Real- und Hauptschuleignung ab:

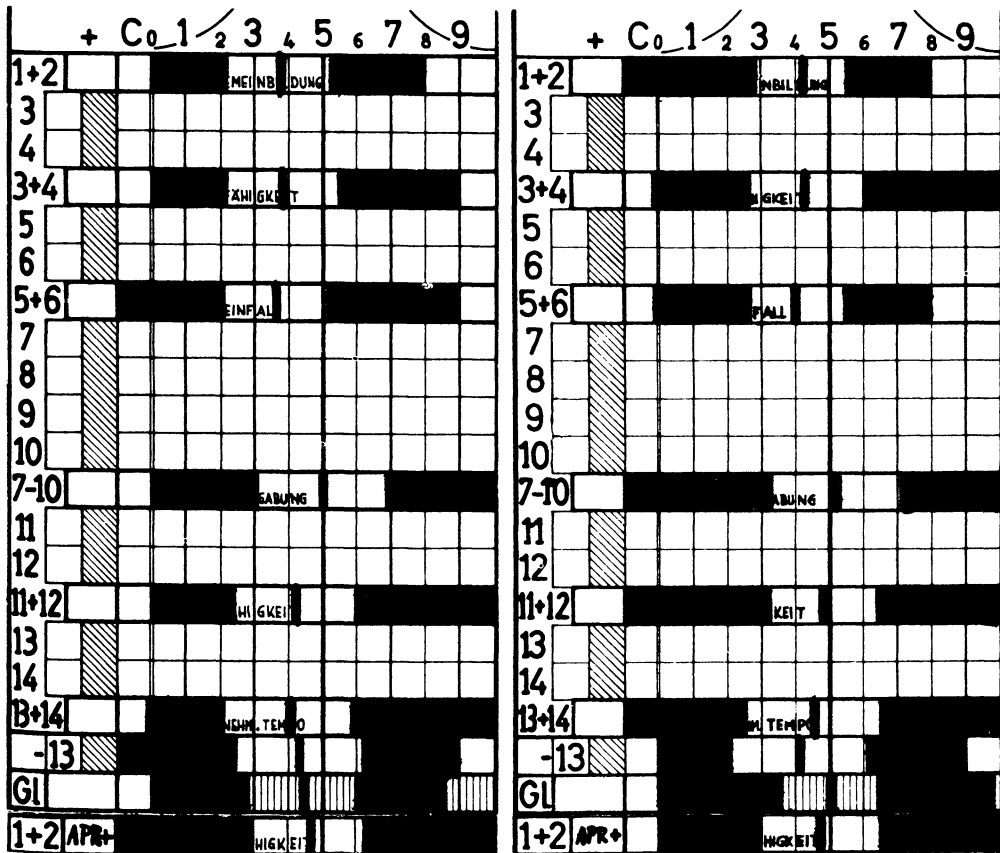
Eine *Gymnasialeignung* manifestiert sich einmal im *höheren intellektuellen Gesamtleistungsniveau* (LPS-GL, PSB-GL, AzN-GL) und zum andern in den *sprachlichen Fähigkeiten* (LPS 1 + 2, 5 + 6, 11 + 12; PSB 1 + 2, 5 + 6; AzN-SE bzw. AN) sowie im *Reasoning-Komplex* (LPS 3 + 4; PSB 3 + 4; AzN-RE, ZR, IV). Hinsichtlich der *technischen* Begaubungsdimensionen (LPS 7-10; PSB 7 + 8; MTVT) treten entsprechende Testleistungsdifferenzen nicht so deutlich in Erscheinung; vor allem verschimmen sich hier zwischen den G- und R-Eignungsgruppen teilweise die Unterschiede. Die Schwerpunkte der *Hauptschuleignung* liegen einmal in den *technischen* Fähigkeitsbereichen (LPS 7-10; PSB 7 + 8; MTVT) resp. mehr *nonverbalen Denkformen* (LPS 3 + 4; PSB 3 + 4) und zum andern in bestimmten *Stützfunktionen* der Intelligenz (LPS-APR, -13; PSB 9 + 10). Die *Realschuleignung* schließlich ist durch eine gewisse Mittelstellung zwischen den aufgewiesenen G- und H-Merkmalsausprägungen gekennzeichnet, wenngleich die R-Eignungskurve eher zum G-Eignungsprofil tendiert. Am wenigsten oder überhaupt nicht unterscheiden sich die drei Schuleignungsgruppen in bezug auf allgemeine Faktoren des Arbeitsverhaltens, der Anstrengungsbereitschaft u.ä. Diese knappen Interpretationshinweise mögen hier genügen.

**Anmerkung zu den nachstehenden Grafiken 5 bis 7 (LPS-Diagrammen):**

Die nicht geschwärzten (ausgesparten) Felder um den arithmetischen Gruppen-Mittelwert (M = senkrechter Balken) bezeichnen die mittlere Dispersion, d.h. den  $\pm 1$  Sigabereich, in dem jeweils rd. 68% aller H- oder R- oder G-Eignungsfälle (der AUKL-Gruppen) im LPS liegen. In den geschwärzten Feldern darunter versus darüber (im LPS-Diagramm links versus rechts vom Mittelwert bzw. Sigabereich) liegen jeweils rd. 16% aller H-, R- oder G-Eignungen. Siehe auch die Erläuterungsskizze zu Grafik Nr. 3 oben.

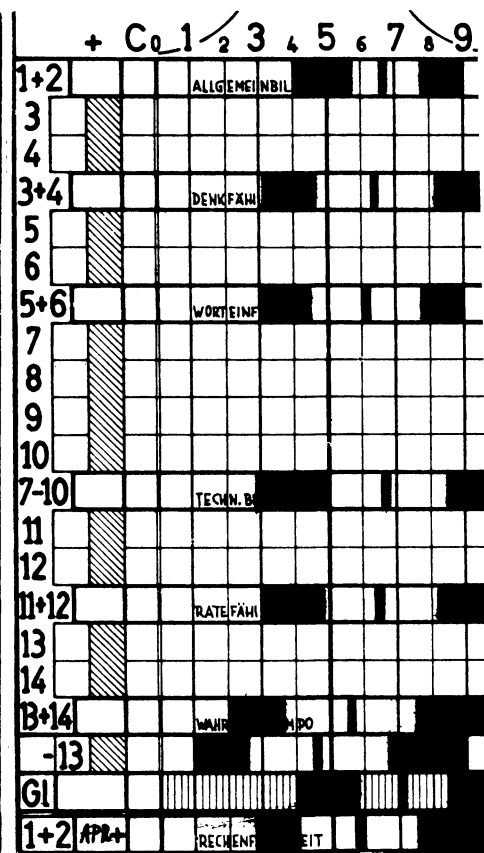
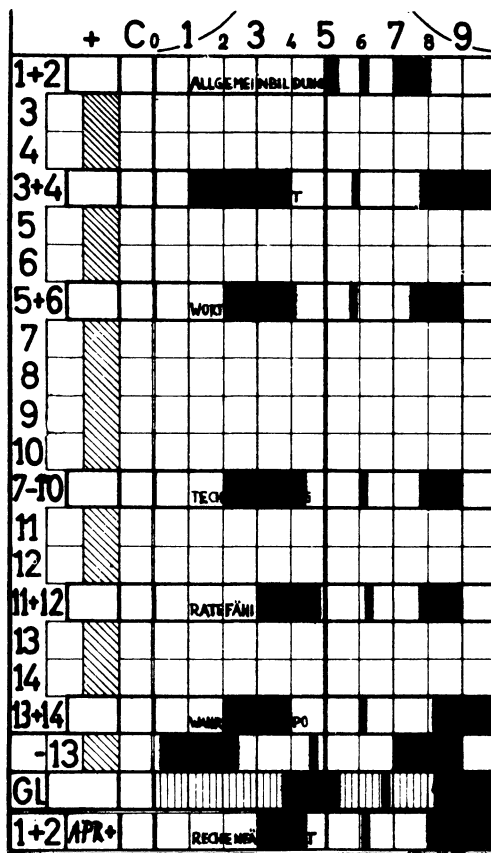
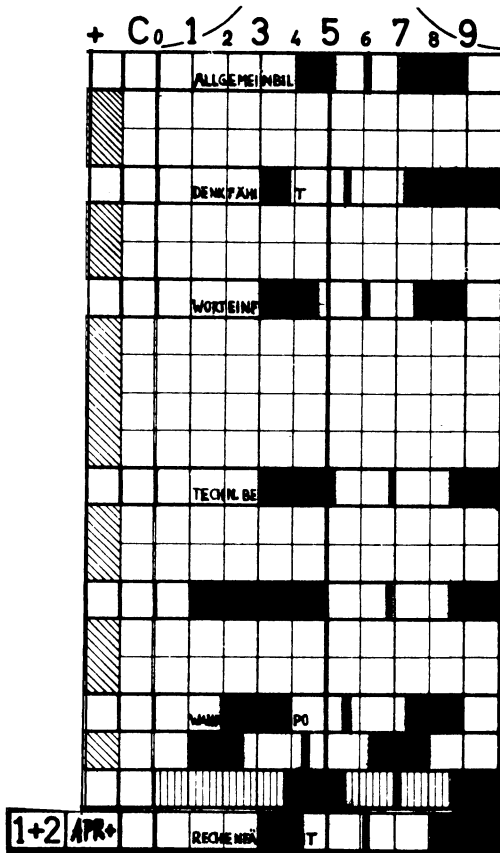
Grafik 5

Eignungskennbereiche im LPS für *Hauptschul*begabungen (H-Klassen 6 und 8)



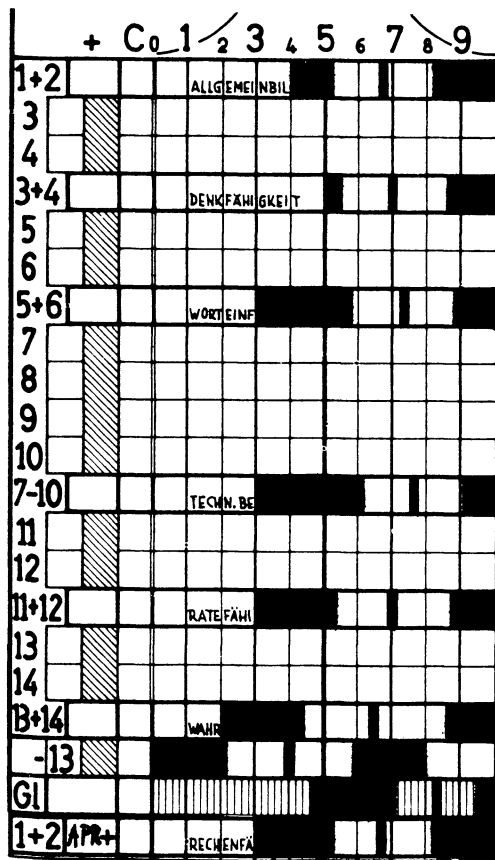
Grafik 5 a  
Eignungskennbereiche  
im LPS für H-Klasse 6

Grafik 5 b  
Eignungskennbereiche  
im LPS für H-Klasse 8

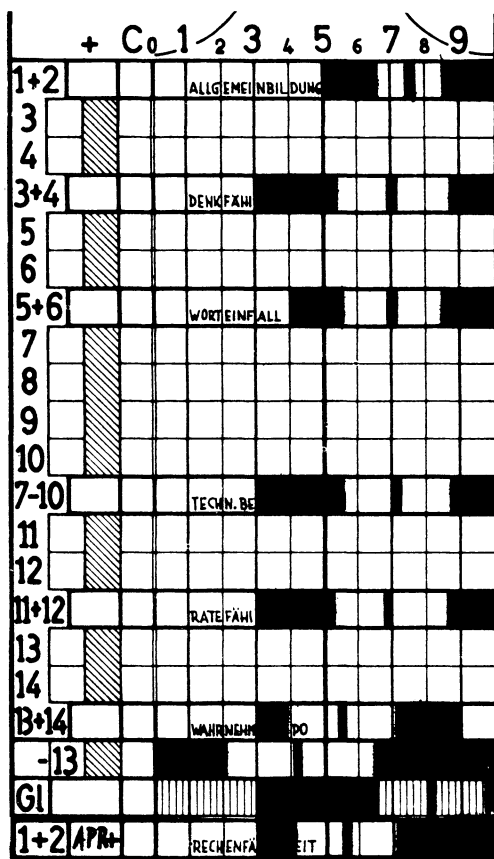


Grafik 6  
Eignungskennbereiche im LPS für *Realschul*begabungen (R-Klassen 6 bis 10)

Grafik 7  
Eignungskennbereiche im LPS für *Gymnasial*begabungen (G-Klassen 6 bis 12)

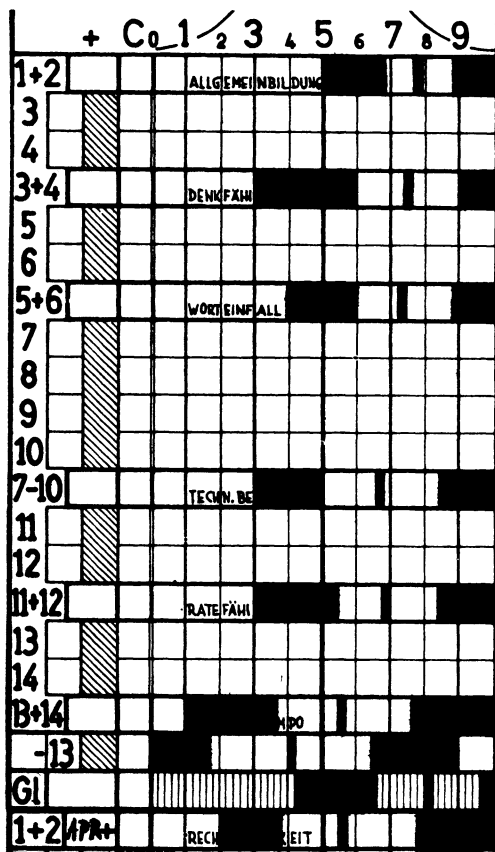


Grafik 7 a  
Eignungskennbereiche  
im LPS für G-Klasse 6

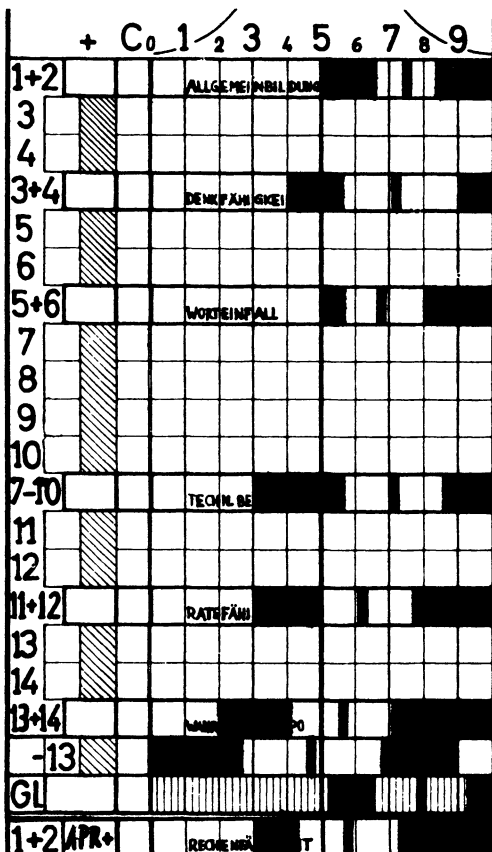


Grafik 7 b  
Eignungskennbereiche  
im LPS für G-Klasse 8





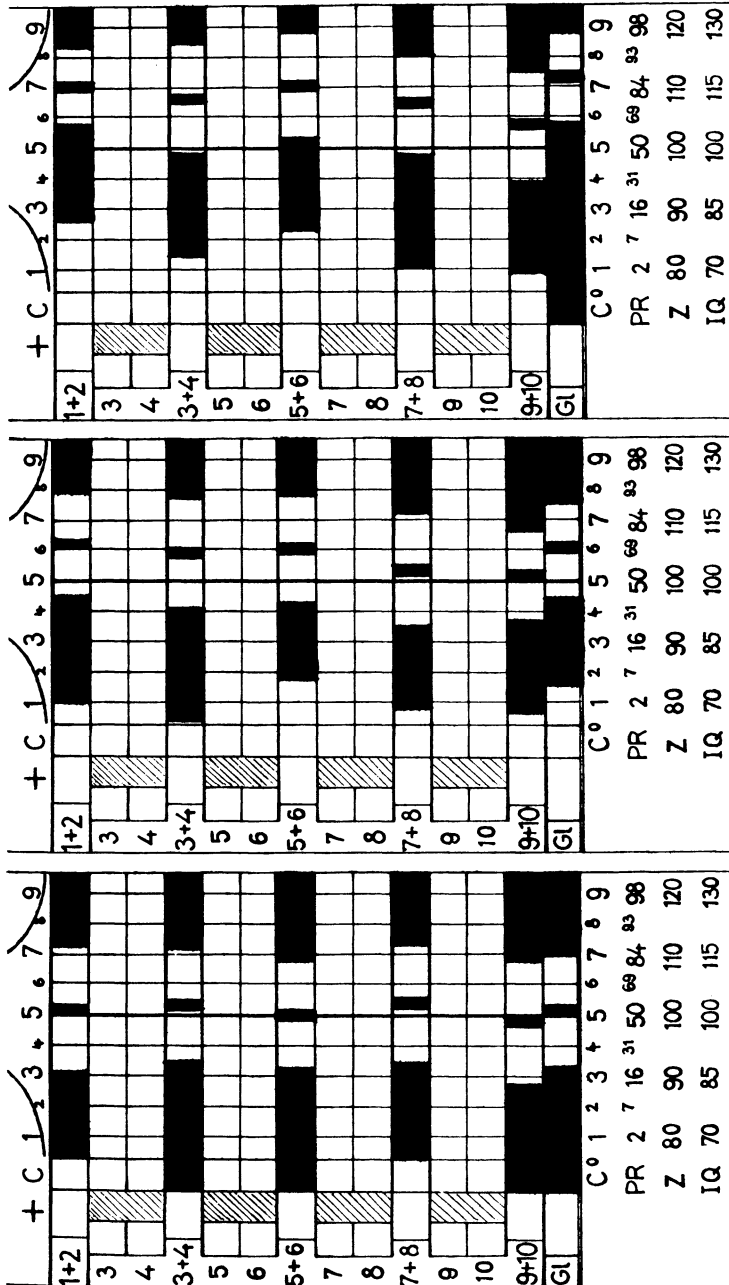
Grafik 7 c  
Eignungskennbereiche  
im LPS für G-Klasse 10



Grafik 7 d  
Eignungskennbereiche  
im LPS für G-Klasse 12

**Anmerkung:** Die nicht geschwärtzten (ausgesparten) Felder um den arithmetischen Gruppen-Mittelwert ( $M$  = senkrechter Balken) bezeichnen die mittlere Dispersion, d.h. den  $\pm 1$  Sigma-Bereich, in dem jeweils rd. 68% aller Hauptschüler oder Realschüler oder Gymnasiasten (der empirischen Schulgruppen) im PSB liegen. In den geschwärtzten Feldern darunter versus darüber (im PSB-Diagramm links versus rechts vom Mittelwert bzw. Sigma-Bereich) liegen jeweils rd. 16% aller Hauptschüler, Realschüler oder Gymnasiasten. Vergleiche auch die Stabdiagramme in Grafik 4 oben.

Grafik 8  
Eignungskennbereiche im PSB für Hauptschüler, Realschüler und Gymnasiasten (empirische Gruppenwerte der Schj./Kl. 5)



Grafik 8 a  
Eignungskennbereiche  
im PSB für H-Klasse 5

Grafik 8 b  
Eignungskennbereiche  
im PSB für R-Klasse 5

Grafik 8 c  
Eignungskennbereiche  
im PSB für G-Klasse 5

Wichtiger für die *eignungsdiagnostische Praxis* sind jedoch die quantitativen Eignungskriterien, also zuverlässige und valide Testprädiktoren. Einer Anregung folgend<sup>53</sup> wurden die wichtigsten Testleistungs- bzw. Eignungskennbereiche für Gymnasium, Realschule und Hauptschule in die originalen LPS- und PSB-Auswertungsdiagramme eingezeichnet. Die LPS-Informationen in Graf. 5 bis 7 beziehen sich hierbei ausschließlich auf die AUKL-differenzierten ("reinen") Schuleignungsgruppen, die PSB-Informationen (Graf. 8) auf die empirischen, also nicht nach AUKL differenzierten G-, R- und H- Gruppen.

Die Benutzung vorstehender Testdiagrammskizzen erleichtert die praktische Schuleignungsfindung in mancherlei Hinsicht. So wird das mühevollen Lesen von Stabdiagrammen ganzer Testbatterien für den Interpreten erheblich vereinfacht; für öfteren Gebrauch empfiehlt sich die Anfertigung von Transparentfolien (mittels Thermo-/Photokopiergerät), die dann als Interpretationsschablonen auf die LPS- und PSB-Diagramme der Originalprotokolle aufgelegt werden können und so einen schnelleren Überblick ermöglichen. Zum andern gestattet die Verwendung unserer Diagrammzeichnungen die Berücksichtigung variabler Anforderungsmaßstäbe, d.h. konkret: Eignungsmaßstäbe für 5. bzw. 6., 8., 10. ("Mittlere Reife") oder 12. ("Prima-Reife") Klassen unterschiedlicher Bildungsformen (Schultypen). Damit wird die Erarbeitung von Schuleignungsprognosen für den Grundschullehrer, den Lehrer an weiterführenden schulischen Bildungseinrichtungen bzw. den Schullaufbahn- oder Bildungsberater erheblich vereinfacht, wenngleich vor einer allzu schematischen Handhabung auch hier gewarnt werden soll.

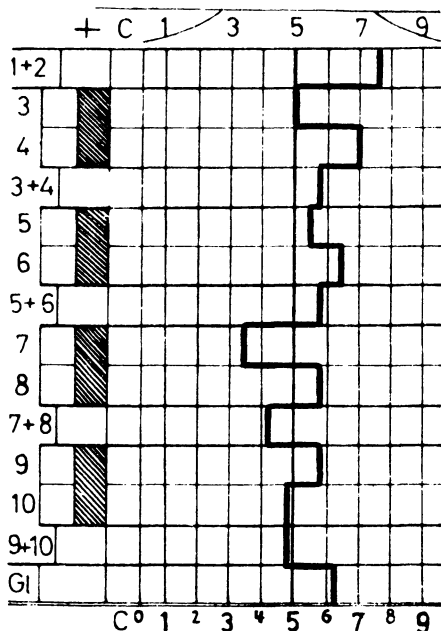
Im allgemeinen wird man davon ausgehen können, daß eindeutige Gymnasial-, Real- oder Hauptschuleignungen jeweils in den *mittleren* (G- R- oder H-) Eignungskennbereichen *oder darüber* (in den Schaubildern rechts vom Betrachter aus) liegen, d.h. Testwerte aufweisen, die im Bereich von -1 Sigma bis + 1 Sigma bzw. + 3 Sigma liegen; in dem so markierten Bereich liegen jeweils 68% plus 16%, also rd. 84% aller betr. G-, R- oder H-Eignungsfälle. Probanden, deren Testwerte *unter* dem -1 Sigmapunkt liegen, kommen in der Regel nicht für die betr. Schuleignungskategorie in Frage; in diesem (unteren) Bereich liegen jeweils nur rd. 16% aller Fälle. Einige *Fallbeispiele* mögen das praktische Vorgehen illustrieren. Zu jedem Fall finden sich folgende Angaben: Name und Alter d. Pb., Jahr d. Schuleignungsuntersuchung, Konf., z.Z. d. Untersuchg. besuchte Schule/Klasse, Wohnortsklasse, Zahl d. Geschw., Beruf und Schulbildg. des Vaters u. der Mutter, Schulzensuren, Testergebnisse (zum Vergleich der in unterschiedlichen Maßskalen ausgedrückten Testergebnisse bzw. Zensuren sei die Benutzung unserer im Anhang beigefügten Transformationstabelle (Tab. VIII) empfohlen), Eignungskategorienschema der Bildungsempfehlungen (LU = Lehrerurteil, TU = Testurteil, BEAE = Bildungsempfehlung nach Approximativer Expertenabstimmung, AUKL = Schullaufbahnempfehlung nach AUKL): GG = Gymnasium geeignet, GB = Gymnasium bedingt geeignet, AG = geeignet für das Aufbau-Gymnasium, RG = Realschule geeignet, RB = Realschule bedingt geeignet, RE = Realschul-Entwicklungsfall, HA = geeignet für Hauptschule A-Kurs, HB = geeignet für Hauptschule B-Kurs, SO = Sonderschule (Verdacht auf Lernbehinderung).

Der "leichteste" Fall ist wohl Nr. 2. Frank ist nach übereinstimmenden Urteilen zur Bildungsempfehlung eindeutig gymnasialgeeignet. Er weist in den für diese Schuleignungsgruppe charakteristischen Testdimensionen im PSB (1 + 2, 3 + 4 5 + 6, GL) sowie im AzN und WST hervorragende Leistungen auf. Ebenso sind seine Schulzensuren in den Hauptfächern durchweg gut. Im Hinblick auf die intellektuellen Schuleignungsvoraussetzungen sowie die aktuellen Schulleistungsverhältnisse des Pb kann eine Gymnasialbildung uneingeschränkt empfohlen werden.

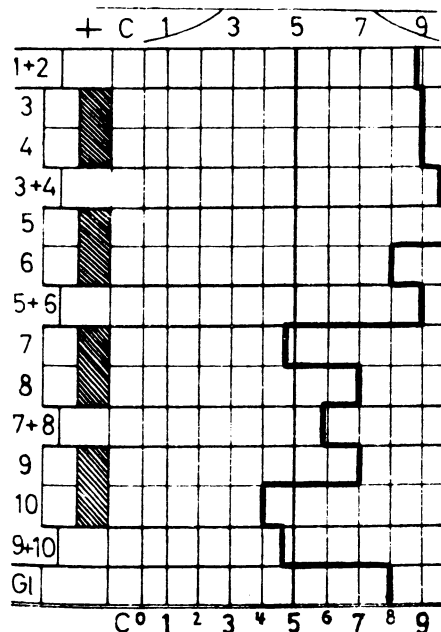
Schwieriger ist schon der Fall Nr. 1. Insgesamt kann man jedoch den TU-, BEAE- und AUKL-Resultaten zustimmen und – entgegen dem Grundschullehrergutachten – eine Realschulbildung vorschlagen. Daß der Pb im AzN relativ schwächere Leistungen zeigt,

<sup>53</sup> Herrn Realschuldirektor R. Wagner, Weinheim, möchte ich dafür auch an dieser Stelle danken.

# PSB-Horn



# PSB-Horn



1. Thomas S.; 10; 11 (1968); o.K.;  
Grundsch.-Kl.4; Großstadt; Geschwi-  
ster: 6; Vaterberuf: einf. Beamter –  
Vo.; Mutterberuf: Facharbeiterin –  
Vo.

*Zensuren:* Aufsatz = 3, Rechtschr. =  
2, Rechn. = 3

*PSB (Horn):* 1 + 2 = 7,6 C

3 + 4 = 5,8

5 + 6 = 5,8

7 + 8 = 4,2

9 + 10 = 4,8

GL = 6,2

*WST 5-6:* GL = 86 PR

*AzN 4+:* GL = 24 PR

RE = 5 RP

AN = 9

ZR = 10

SE = 7

IV = 8

*Bildungsempfehlung:* LU = HA

TU = RB

BEAE = RG

AUKL = RG

2. Frank H.; 9; 4 (1968); ev.; Grundsch.-  
Kl. 4; Großstadt; Geschwister: 1;  
Vaterberuf: mittl. Angest. – MR;  
Mutterberuf: Hausfrau – Vo.

*Zensuren:* Aufsatz = 2, Rechtschr. =  
2, Rechn. = 2

*PSB (Horn):* 1 + 2 = 8,8 C

3 + 4 = 9,8

5 + 6 = 9,0

7 + 8 = 5,8

9 + 10 = 4,6

GL = 8,0

*WST 5-6:* GL = 92 PR

*AzN 4+:* GL = 96 PR

RE = 7 RP

AN = 19

ZR = 16

SE = 10

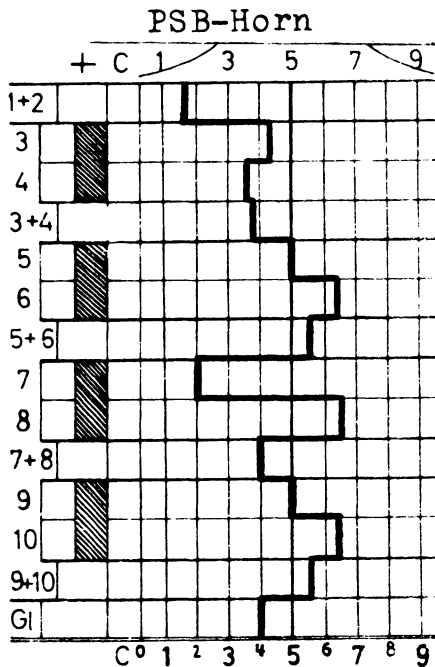
IV = 10

*Bildungsempfehlung:* LU = GG

TU = GG

BEAE = GG

AUKL = GG



3. Anneliese T.; 10; 1 (1968); rk.; Grundschul-Kl. 4; Großstadt; Geschwister: 5; Vaterberuf: Ing. – MR.; Mutterberuf: Hausfrau – Vo.

**Zensuren:** Aufsatz = 3, Rechtschr. = 3, Rechn. = 3

**PSB (Horn):** 1 + 2 = 1,6 C  
 3 + 4 = 3,8  
 5 + 6 = 5,6  
 7 + 8 = 4,0  
 9 + 10 = 5,6  
 GL = 4,0

**WST 5-6:** GL = 38 PR

**AzN 4+:** GL = 38 PR  
 RE = 6 RP  
 AN = 15  
 ZR = 7  
 SE = 7  
 IV = 8

**Bildungsempfehlung:** LU = GG  
 TU = HA  
 BEAE = HA  
 AUKL = HA

geht wohl in erster Linie zu Lasten der Kurzschuljahre. Der AzN ist ein Verfahren, das sich stärker schulleistungsrelevant und damit von Lernerfahrungen abhängig erweist, so daß sich verkürzte Beschulungsdauer bzw. reduzierte Lernmöglichkeiten gerade hier bemerkbar machen müssen, viel stärker jedenfalls als beim lernleistungsunabhängigeren PSB. Hinzu kommt, daß im PSB *Altersnormen*, im AzN jedoch (bis 1967) nur *Schul-* bzw. *Klassennormen* vorlagen<sup>54</sup>.

Zweifelloos der interessanteste und problematischste Fall ist im Beispiel Nr. 3 gegeben. Sofort fällt hier die Diskrepanz von LU und den übrigen Urteilsinstanzen bezüglich der Schullaufbahneempfehlung auf. Die Vermutung liegt nahe, daß hier viel stärker als in den beiden anderen Fallbeispielen das Lehrerurteil durch den Bildungswunsch der Pb-Eltern (Pb A. trat zu Beginn des Schuljahrs 1968/69 ins Gymnasium über) beeinflusst wurde. Sofern man dem PSB-Leistungsprofil (allein) folgt, wäre kaum mehr als Hauptschuleignung zu testieren. Andererseits verfügt die Pb über eine relativ gute (interindividuell verglichen jedoch nur durchschnittliche) AzN-Leistung. Ich selbst hätte im vorliegenden Fall eine RE-Diagnose gestellt. Der sehr unausgeglichene Kurvenverlauf im PSB-Profil deutet in den Spitzen (PSB 6, 8, 10) an, wo in etwa das *potentielle* (allgemeine) Intelligenzniveaus der Pb liegen könnte. Nicht selten steckt hinter solchen diskrepanten Leistungsbildern ein bestimmter Problemkomplex, z.B. persönlichkeitspsychologischer oder sozialer Art. In jedem Falle wäre der Pb A. unverzüglich eine schulpyschologische Beratung – möglichst unter Beteiligung der Eltern – anzuraten.

Die im folgenden angeführten Beispiele praktischer Fallbeurteilung sind vor allem für Übungszwecke ausgewählt worden. Hierbei wurde nicht nur auf eine breite Streuung der Pbn unterschiedlicher Provenienz, sondern auch auf die Variation der Testauswahl

<sup>54</sup> In der 1967 erschienenen Neuauflage des AzN liegen nunmehr (unter Berücksichtigung der Kurzschuljahre) auch Altersnormen vor (vgl. S. 130 oben). Die AzN-Daten unserer Fallbeispiele basieren ausnahmslos noch auf den (älteren) Schul-/Klassennormen.

geachtet. Schließlich sind – um das Interesse der Leser nicht zu beeinträchtigen – hinfür die seinerzeit ausgesprochenen Bildungsempfehlungen weggelassen worden. Diese können (zur Kontrolle) am Ende unserer Beispielsammlung in der Fußnote 55 eingesehen werden.

4. Rainer O.; 10; 10 (1968); rk.; Grundschul-Kl. 4 (Repetent); Großstadt; Geschwister: 1; Vaterberuf: Kraftfahrer – Vo.; Mutterberuf: Hausfrau – Vo.

*Zensuren:* Aufsatz = 4, Rechtschr. = 5, Rechnen = 3

*PSB (Horn):* 1 + 2 = 1,4 C

3 = 6,2

4 = 6,6

3 + 4 = 6,2

5 = 3,4

6 = 3,0

5 + 6 = 2,8

7 = 3,8

8 = 5,8

7 + 8 = 4,4

9 = 1,0

10 = 5,4

9 + 10 = 3,4

GL = 3,0

*WST 5-6:* GL = 18 PR

*AzN 4+:* GL = 13 PR

RE = 2 RP

AN = 13

ZR = 7

SE = 4

IV = 7

5. Karin F.; 12; 10 (1965); rk.; H-Kl. 6; Landgemeinde; Geschwister: 1; Vaterberuf: Gipser – Vo.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch: Näherin

*Zensuren:* Lesen = 2, Aufsatz = 3, Sprachkunde = 3, Rechtschr. = 3, Rechnen = 2/3

*LPS (Horn):* 1 + 2 = 5 C

3 + 4 = 7

5 + 6 = 5

7 – 10 = 6

11 + 12 = 4

13 + 14 = 4

- 13 = 6

GL = 5,4

APR = 4

*WST 5-6:* GL = 96 PR

*MTVT:* GL = 20 PR

*BIT (Irle):* TH = 55 PR

GH = 85

TN = 30

EH = 75

LF = 50

KB = 60

VB = 35

LG = 68

SE = 56

6. Hans R.; 12; 9 (1965); rk.; H-Kl. 6; Kleinstadt; Geschwister: 1; Vaterberuf: Werkmeister – Vo.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch: Laborant

*Zensuren:* Lesen = 2, Aufsatz = 3, Sprachkunde = 3, Rechtschr. = 2, Rechnen = 2

*LPS (Horn):* 1 + 2 = 7 C

3 + 4 = 6

5 + 6 = 6

7 – 10 = 5

11 + 12 = 7

13 + 14 = 3

- 13 = 3

GL = 6,4

APR = 7

*WST 5-6:* GL = 98 PR

*MTVT:* GL = 50 PR

*BIT (Irle):* TH = 40 PR

GH = 55

TN = 60

EH = 35

LF = 65

KB = 55

VB = 25

LG = 90

SE = 45

7. Renate M.; 12; 7 (1965); ev.; H-Kl. 6; Kleinstadt; Geschwister: 2; Vaterberuf: Maurer (Polier) – Vo.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch: Friseur

**Zensuren:** Lesen = 2, Aufsatz = 3, Sprachkunde = 3, Rechtschreib. = 2, Rechnen = 3

<b>LPS (Horn):</b>	1 + 2 = 7 C	<b>WST 5-6:</b> GL = 98 PR	<b>BIT (Irle):</b> TH = 10 PR
	3 + 4 = 8		GH = 94
	5 + 6 = 7	<b>MTVT:</b> GL = 25 PR	TN = 7
	7 – 10 = 9		EH = 50
	11 + 12 = 9		LF = 72
	13 + 14 = 6		KB = 70
	- 13 = 8		VB = 45
	GL = 9,6		LG = 70
	APR = 7		SE = 70

8. Waltraud K.; 12; 3 (1965); rk.; H-Kl. 6; Landgemeinde; Geschwister: 1; Vaterberuf: Blechner – Vo.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch: Verkäuferin

**Zensuren:** Lesen = 3, Aufsatz = 3, Sprachkunde = 3, Rechtschr. = 4, Rechnen = 3

<b>LPS (Horn):</b>	1 + 2 = 5 C	<b>WST 5-6:</b> GL = 62 PR	<b>BIT (Irle):</b> TH = 28 PR
	3 + 4 = 5		GH = 67
	5 + 6 = 2	<b>MTVT:</b> GL = 5 PR	TN = 22
	7 – 10 = 4		EH = 72
	11 + 12 = 3		LF = 50
	13 + 14 = 4		KB = 85
	- 13 = 4		VB = 50
	GL = 4,2		LG = 70
	APR = 4		SE = 72

9. Monika S.; 12; 5 (1965); ev.; H-Kl. 6; Kleinstadt; Geschwister: 1; Vaterberuf: Bauing. – Mi.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch: ?

**Zensuren:** Lesen = 2, Aufsatz = 2, Sprachkunde = 2, Rechtschr. = 2, Englisch (Wahlfach) = 2  
Rechnen = 2

<b>LPS (Horn):</b>	1 + 2 = 6 C	<b>WST 5-6:</b> 96 PR	<b>BIT (Irle):</b> TH = 38 PR
	3 + 4 = 5		GH = 98,8
	5 + 6 = 7	<b>MTVT:</b> 42 PR	TN = 5
	7 – 10 = 7		EH = 72
	11 + 12 = 6		LF = 73
	13 + 14 = 5		KB = 30
	- 13 = 6		VB = 30
	GL = 7,4		LG = 42
	APR = 4		SE = 40

10. Peter K.; 11; 11 (1965); rk.; H-Kl. 6; Landgemeinde; Geschwister: 2; Vaterberuf: Müllermeister – Vo.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch: Müller

**Zensuren:** Lesen = 2, Aufsatz = 2, Sprachkunde = 2, Rechtschr. = 2, Rechnen = 2

<b>LPS (Horn):</b>	1 + 2 = 7 C	<b>WST 5-6:</b> 90 PR	<b>BIT (Irle):</b> TH = 60 PR
	3 + 4 = 9		GH = 90
	5 + 6 = 6	<b>MTVT:</b> 94 PR	TN = 25
	7 – 10 = 10		EH = 70
	11 + 12 = 6		LF = 50
	13 + 14 = 7		KB = 70
	- 13 = 5		VB = 43
	GL = 8,8		LG = 42
	APR = 6		SE = 50

11. Ingrid B.; 15; 11 (1965); ev.; Kl. UII (10) neuspr. Gymn.; Kleinst./Mittelstadt;  
Geschwisterzahl: 1; Vaterberuf: Kfm. – Mi.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch:  
Lehrerin

*Zensuren:* Deutsch = 3, Latein = 5, Engl. = 4, Franz. = 3, Math. = 5

<i>LPS (Horn):</i>	1 + 2 = 7 C	<i>MTVT:</i>	GL = 40 PR	<i>BIT (Irle):</i>	TH = 15 PR
	3 + 4 = 5				GH = 92
	5 + 6 = 5				TN = 15
	7 – 10 = 4				EH = 70
	11 + 12 = 3				LF = 35
	13 + 14 = 3				KB = 65
	- 13 = 8				VB = 72
	GL = 5,2				LG = 52
	APR = 6				SE = 80

12. Erika A.; 16; 11 (1966); rk.; Kl. UII (10) neuspr. Gymn.; Klein/Mittelstadt; Geschwisterzahl: 1; Vaterberuf: Elektromeister – Vo.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch: Dolmetscherin

*Zensuren:* Deutsch = 2, Latein = 3, Engl. = 2, Franz. = 2, Math. = 3

<i>LPS (Horn):</i>	1 + 2 = 6 C	<i>MTVT:</i>	GL = 50 PR	<i>BIT (Irle):</i>	TH = 1 PR
	3 + 4 = 5				GH = 55
	5 + 6 = 5				TN = 70
	7 – 10 = 4				EH = 4
	11 + 12 = 4				LF = 2
	13 + 14 = 7				KB = 90
	- 13 = 6				VB = 77
	GL = 5,4				LG = 96
	APR = 4				SE = 2

13. Wolfgang F.; 18; 10 (1965); rk.; Kl. UII (10); math.-nat. Gymn.; Großstadt; Rep. in Kl. Vu. UII; Geschwisterzahl: 0; Vaterberuf: Dipl.-Chem. – Hochsch.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch: Chemiker

*Zensuren:* Deutsch = 3, Engl. = 6, Latein = 6, Math. = 4, Phys. = 3, Chemie = 4, Biol. = 2, Gesch. = 3, Musik = 4, Leibeserz. = 3

<i>LPS (Horn):</i>	1 + 2 = 5 C	<i>KLT (Lienert):</i>	Leistg. = 29 PR
	3 + 4 = 3		Fehler/Arbeitsqualität = 1,0 FQ ( $\phi$ )
	5 + 6 = 4	<i>PIT (Mittenecker/Toman):</i>	Mang. Selbstkritik und
	7 – 10 = 3		Extraversion stark ausgeprägt versus unter-
	11 + 12 = 4		durchschnittl. Selbstkritik und Introversion;
	13 + 14 = 3		überdurchschnittl. hohe Ausprägung der Inter-
	- 13 = 7		essen für Wissenschaft, Umgang mit Menschen
	GL = 4,2		in Geschäft und Wirtschaft, sowie soz. Berufe
	APR = 5		(Lehrer- und Erzieherberufe etc.); schwache
			Interessen für Verrechng. und Verwaltung
			sowie Bild. Kunst.
	TH = 67 PR		
	GH = 12		
	TN = 95		
	EH = 60		
	LF = 30		
	KB = 38		
	VB = 40		
	LG = 85		
	SE = 15		

Zu den Fallbeispielen Nr. 4 bis 15 wurden seinerzeit folgende Schuleignungsurteile abgegeben: Nr. 4 = HB, Nr. 5 = RE, Nr. 6 = RG, Nr. 7 = GG, Nr. 8 = HA, Nr. 9 = RG (AG), Nr. 10 = GG, Nr. 11 = RB, Nr. 12 = RG (?), Nr. 13 = HA, Nr. 14 = GG, Nr. 15 = GG. In einigen Fällen (z.B. Nr. 15 oder auch Nr. 14) wären besondere pädagogisch-psychologische Maßnahmen indiziert. Bei Fall Nr. 12 ist das Testergebnis vermutlich irreliabel, worüber jedoch eine Retestung Aufschluß geben könnte.



14. Birgit H.; 16; 7 (1965); ev.; Kl. UII (10) neuspr. Gymn.; Großstadt; Vaterberuf: Lehrer – Hochsch.; Mutterberuf: Sekretärin – Mi.; Geschwister: 0; Pb-Berufswunsch: Architektin, Jugendrichterin

**Zensuren:** Deutsch = 3, Engl. = 4, Latein = 4, Franz. = 4, Math. = 4, Phys. = 3, Chemie = 4, Biol. = 3, Gesch. = 3, Musik = 4, Sport = 3

**LPS (Horn):** 1 + 2 = 9 C  
 3 + 4 = 8  
 5 + 6 = 8  
 7 – 10 = 8  
 11 + 12 = 10  
 13 + 14 = 8  
 - 13 = 3  
 GL = 10,6  
 APR = 7

**KLT (Lienert):** Leistg. = 94 PR  
 Fehler/Arbeitsqualität = 0,46 FQ (unt.  $\phi$ )

**PIT (Mittenecker/Toman):** Mang. Selbstkritik und Introversion unterdurchschnittl. versus Selbstkritik und Extraversion positiv (stark) ausgeprägt; überdurchschnittl. Interessen für Handwerk, Bild. Kunst. und soz. Berufe, schwache Interessen für Musik

**BIT (Irle):** TH = 15 PR  
 GH = 97,5  
 TN = 50  
 EH = 30  
 LF = 10  
 KB = 2  
 VB = 10  
 LG = 92,5  
 SE = 85

15. Florian S.; 15; 3 (1965); Kl. UII (10) altspr. Gymn.; Großstadt; Vaterberuf: Syndikus Hochsch.; Mutterberuf: Hausfrau – Vo.; Pb-Berufswunsch: Jurist

**Zensuren:** Deutsch = 4, Latein = 3, Engl. = 5, Griech. = 5, Math. = 2, Phys. = 2, Chem. = 4, Biol. = 4, Gesch. = 4, Musik = 2

**LPS (Horn):** 1 + 2 = 9 C  
 3 + 4 = 10  
 5 + 6 = 10  
 7 – 10 = 8  
 11 + 12 = 8  
 13 + 14 = 7  
 - 13 = 0  
 GL = 10,4  
 APR = 8

**KLT:** Leistg. = 92 PR, Qual. = 0,5 FQ (unt.  $\phi$ )

**Aus PIT + BIT:** Ausgeprägte (starke) Int. für GH, EH u. LG bzw. Bild. Kunst. u. Musik. Schwache Interessen für SE und Wissensch. Mangelnde Selbstkritik (erheblich vom Durchschnitt abweichend)

### Exkurs: Eignungskriterien für den Englischunterricht an der Hauptschule (Hauptschule A-Kurs)

Für alle diejenigen Viertklässer der Grundschule, die nicht auf die Realschule oder das Gymnasium übertreten, stellt sich beim Eintritt in die Hauptschule die Frage, ob sie am Englischunterricht (Hauptschule A-Kurs) teilnehmen sollen oder nicht, d.h. ob von ihren Begabungsvoraussetzungen her Erfolgchancen bestehen. Oder anders formuliert: Welche Prädiktorvariablen sagen etwas über die erfolgreiche Teilnahme am Englischunterricht der Hauptschule aus? Dieser Frage ging U. Beck in einer vom Verfasser betreuten Arbeit nach. Die von Beck (1969) untersuchte *Stichprobe* – 50 Volksschüler der Schj.-Klasse 5 – wurde nach Geschlecht, G/R-Übertrittsquote, Sozialgliederung (sozio-ökonomischer Status variable), Wohngebiet u.a. relevanten Kriterien kontrolliert und kann in etwa als repräsentativ für klein- und mittelstädtische Schulverhältnisse in Baden-Württemberg gelten. Als *Untersuchungsverfahren* wurden Tests (AzN 4+, FTU 4-6, PSB) und Grundschulzensuren bzw. sog. Probearbeiten in der 4. Grundschulklasse verwendet. Als *Validitätskriterien* dienten die Lehrerurteile über Fremdspracheneignung (Einteilung in A- und B-Kursteilnehmer: nur die A-Kursteilnehmer (N = 28) beteiligten sich am Englischunterricht) und

der Unterrichtserfolg in Englisch (Englisch-Zensuren nach 1 Jahr Unterricht an der Hauptschule). Die *Ergebnisse der statistischen Analyse* lassen sich folgendermaßen zusammenfassen:

1. Die Knaben und Mädchen der Untersuchungsstichprobe unterschieden sich *sehr signifikant* ( $p$  kleiner 1%) in den "Testdiktaten" 1 u. 2 (Rechtschreib-Probearbeiten) sowie in den Grundschulzensuren in Deutsch bzw. Aufsatz, Rechtschreiben und Lesen. Bezüglich der übrigen Leistungsvariablen (AzN, FTU, PSB; Probearbeiten 1 u. 2 in Rechnen; Zensuren in Sprachkunde, Heimatkunde, Rechnen und Englisch) konnten keine oder nur zufällige geschlechtsspezifische Differenzen nachgewiesen werden. Es bestätigt sich also auch hier die sonst gewonnene Erfahrung, daß Mädchen vor allem in den sprachlichen Schulleistungsdimensionen bessere Zensuren aufweisen (vgl. z.B. Aurin et al. 1968, S. 98 oder Kemmler, 1967, S. 94). Übereinstimmend interpretieren Kemmler, Aurin u.a. diese Befunde im Sinne von Interdependenzen bestimmter Erwartungseinstellungen seitens der Lehrer bzw. der Gesellschaft einerseits und entsprechender Anpassungstendenzen bzw. realisierter Anpassung seitens der weiblichen Schüler andererseits. Analog dazu deutet Beck (1969, S. 81) ihre Ergebnisse: "In der Englischleistung tritt jedoch kein signifikanter Unterschied zwischen Knaben und Mädchen auf. Da auch im AzN 4+ und FTU 4-6 und in den Teilttests des PSB, die den Verbalfaktor erfassen, keine signifikanten Unterschiede auftreten, dürfte der Unterschied zwischen Knaben und Mädchen bei dieser Stichprobe nicht durch echte Sprachbegabungsunterschiede bedingt sein, sondern durch die verschiedene Arbeitshaltung, eventuell auch durch unterschiedliche Leistungsbewertung".
2. Die erfolgreichen Englischschüler (Hauptschüler, die mindestens 1 Jahr lang erfolgreich am Englischunterricht teilnahmen) und die B-Kursteilnehmer (nicht am Englischunterricht beteiligten bzw. dafür geeignet erschienenen Hauptschüler) offenbarten *sehr signifikante* Leistungsunterschiede in den Testvariablen AzN 4+; FTU 4-6; PSB-GL, 5, 5 + 6, 9 + 10; in den Testdiktaten 1 + 2 sowie in den Grundschulzensuren für Deutsch (Aufsatz, Rechtschreiben, Sprachkunde, Lesen) und Heimatkunde resp. *signifikante* ( $p$  größer 1% und kleiner 5%) Unterschiede im PSB 1 + 2 und 10 sowie im Rechnen (Schulzensur und Rechen-Probearbeiten 1 + 2). In den Testleistungen zum PSB 3, 4, 3 + 4, 6, 7, 8, 7 + 8 und 9 unterschieden sich die beiden Hauptschulgruppen nicht signifikant.
3. Unter anderem Gruppierungsgesichtspunkt, nämlich im Leistungsvergleich der ausschließlich nach LU (Klassenlehrer) für den Englischunterricht ausgesuchten versus nicht berücksichtigten Hauptschüler, ergaben sich lediglich *sehr signifikante* Differenzen im AzN sowie *signifikante* Leistungsunterschiede im PSB 5, 10 und 9 + 10. Die Schulzensuren wurden bei diesem Vergleich nicht kontrolliert.
4. Besonders aufschlußreich für die Erkennung der Fremdspracheneignung in der Hauptschule ist eine Gruppengegenüberstellung der "guten" und "schlechten" Englischschüler, d.h. derjenigen Hauptschüler, die Englischzensuren von 2 und 3 versus von 4 und 5 (nach 1 Jahr Englischunterricht) aufwiesen. Bei dieser Gruppendifferenzierung ergaben sich *sehr signifikante* Leistungsunterschiede im PSB 1 + 2 sowie in Deutsch bzw. *signifikante* Differenzen in PSB-GL, 5, 5 + 6, 9 sowie im Rechtschreiben und in Sprachkunde. Keine genügende Differenzierungskraft unter dem genannten Kriterium bewies u.a. hier der FTU, was in gewisser Hinsicht überrascht.
5. Um die wichtigsten Prädiktorvariablen für die Fremdspracheneignung von Hauptschülern zu ermitteln, wurden abschließend sämtliche Test- und anderen Variablen mit der Englischzensur korreliert. Bei konstanter Stichprobengröße ( $N = 28$ ) sind die in Tab. 8 wiedergegebenen Korrelationskoeffizienten  $r_{xy}$  direkt miteinander vergleichbar. Zur Erleichterung der Interpretation wurden die sehr signifikanten (ss) und signifikanten (s) bzw. nicht signifikanten (ns;  $p$  größer 5%) Korrelationskoeffizienten in der letzten Spalte rangmäßig durchnummeriert. Für die abschließende Interpretation sind dabei insbesondere die ersten 10 (ss) Koeffizienten von Bedeutung.

Tabelle 8

Korrelative Zusammenhänge von Englisch-Zensur (Engl.) und verschiedenen Test- bzw. Schulleistungskriterien bei 28 Hauptschülern der 5. Klasse – nach U. Beck (1969, S. 79 f.)

Beziehung zwischen den Kriterien	$r_{xy}$	Signifikanz des r-Wertes	Rangposition des r-Wertes
Engl. – AzN 4+	+ 0.2801	ns	(19.)
Engl. – FTU 4-6	+ 0.4450	s	11.
Engl. – PSB-GL	+ 0.6685	ss	3.
Engl. – PSB 1 + 2	+ 0.6808	ss	2.
Engl. – PSB 3	+ 0.1055	ns	(25.)
Engl. – PSB 4	+ 0.2910	ns	(18.)
Engl. – PSB 3 + 4	+ 0.2078	ns	(22.)
Engl. – PSB 5	+ 0.5270	ss	8.
Engl. – PSB 6	+ 0.0435	ns	(26.)
Engl. – PSB 5 + 6	+ 0.5153	ss	9.
Engl. – PSB 7	+ 0.4777	ss	10.
Engl. – PSB 8	+ 0.1281	ns	(24.)
Engl. – PSB 7 + 8	+ 0.4274	s	13.
Engl. – PSB 9	+ 0.4040	s	14.
Engl. – PSB 10	– 0.0891	ns	(27.)
Engl. – PSB 9 + 10	+ 0.1793	ns	(23.)
Engl. – Aufsatzzensur	+ 0.3359	ns	(17.)
Engl. – Rechtschreibzensur	+ 0.5382	ss	7.
Engl. – Sprachkundezensur	+ 0.4326	s	12.
Engl. – Lesezensur	+ 0.5421	ss	6.
Engl. – Deutschzensur	+ 0.6896	ss	1.
Engl. – Rechenzensur	+ 0.2256	ns	(20.)
Engl. – Heimatkundezensur	+ 0.3429	ns	(15.)
Engl. – Rechn. + Heimatk.zensur	+ 0.3396	ns	(16.)
Engl. – Probearb. Rechn. 1 + 2	+ 0.2116	ns	(21.)
Engl. – Probearb.Rechtschr. 1 + 2	+ 0.5980	ss	4.
Engl. – Probearb.Rechn.+Rechtschr.	+ 0.5527	ss	5.

Als Resümee der Beckschen Untersuchungen können wir somit festhalten: Vor allen anderen *Prädiktoren der Fremdspracheneignung (Englischeignung) in der Hauptschule* kommt der *Deutschzensur* ( $r = 0.69$ ) sowie den Testleistungen im *PSB 1 + 2* ( $r = 0.68$ ) und *PSB-GL* ( $r = 0.67$ ) Prognosegültigkeit zu, gefolgt von den Prädiktoren *Probearbeiten* (Rechtschreiben versus Rechtschreiben + Rechnen), *Lese-* und *Rechtschreibzensuren* sowie den *PSB*-Tests 5, 5 + 6, 7. In geringerem Maße eignet dem FTU 4-6, der Sprachkundezensur, PSB 7 + 8 und PSB 9 eine gewisse Vorhersagegültigkeit; interessanterweise rangiert der FTU 4-6 mit einem Validitätskoeffizienten von  $r = 0.45$  erst an 11. Stelle der aufgezählten Prädiktoren. Daraus folgert Beck abschließend: "Bei der Auslese für den Englischunterricht an einer Hauptschule . . . ist also neben einer Auslese nach Deutschzensuren (und Probearbeit Diktat) diejenige nach Leistungen im PSB 1 + 2, PSB-GL und PSB 5 ratsam. Die Auslese auf Grund der Leistungen im AzN 4+ und FTU 4-6 könnte sich dagegen als nicht befriedigend erweisen, ebenso eine Auslese auf Grund der Schulzensuren in Rechnen und Heimatkunde" (loc. cit., S. 88). Damit beenden wir die Diskussion intelligenzdiagnostischer Probleme in der Hauptschule.

Eine Reihe der im nächsten Abschnitt angesprochenen Diagnostikprobleme wird sicherlich auch für den Grund- und Sekundarstufenlehrer manche relevante Fragenkomplexe behandeln. Die Lektüre der folgenden Seiten sei deshalb nicht nur Sonderschulpädagogen, sondern allen, die sich mit Fragen pädagogisch-psychologischer Schülerbeurteilung befassen wollen oder müssen, empfohlen.

## 2. Intelligenzdiagnostische Probleme in der Sonderpädagogik

### a) Probleme und Ergebnisse der Lernbehindertendiagnostik

Die Lernbehinderten (Lb) stellen mit Abstand die größte Gruppe der Sonderschüler, nach den Angaben von Brackens (s.S. 112 oben) rd. 6% eines unausgelesenen Altersjahrgangs. Im Rahmen unserer baden-württembergischen Begabungsuntersuchungen bei insgesamt 21 399 Viertklässkindern fanden sich 1967/68 immerhin noch 4,3% Lernbehinderte Schüler oder zumindest Lb-Verdachtsfälle laut Testurteil. Allerdings variierten die Lb-Quoten regional sehr stark. Während in städtischen Wohnregionen – erwartungsgemäß – die niedrigsten Quoten, nämlich durchschnittlich 3,4% (bei einer Min.-Quote von 0,7% und einer Max.-Quote von 6,3%) beobachtet wurden, streuten die Lb-Quoten auf dem Lande zwischen 9,0% und 1,6% und lagen im Mittel bei 5,3%, also nur knapp unter dem von Bracken geschätzten Wert. Dabei ist zu berücksichtigen, daß es sich bei den von uns angegebenen Werten um Quoten bereits (teilweise) ausgelesener Schuljahrgänge handelt, da lernbehinderte Kinder in der Regel schon vor dem 4. Schuljahr aus der Grundschule (Regelschule) herausgenommen werden.

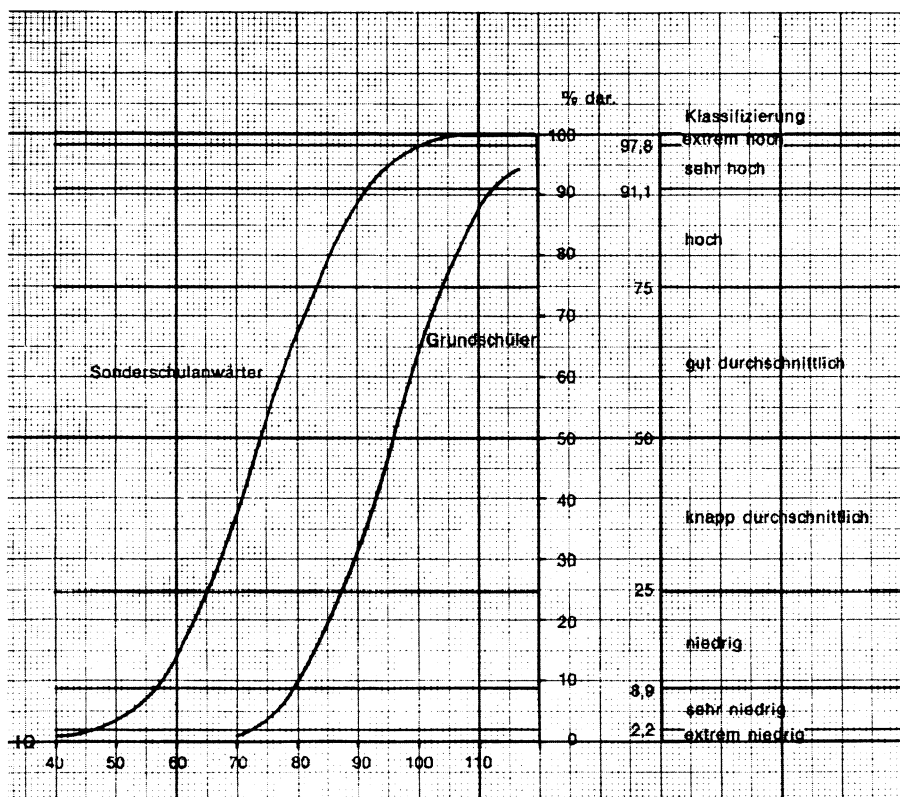
Über die bei Lernbehinderten einschlägigen psychodiagnostischen Untersuchungsverfahren liegt ein umfassendes Kompendium von Zimmermann (1969) vor, so daß wir uns auf einige zentrale intelligenzdiagnostische Problemfragen hier beschränken können. Als Instrument zur Ermittlung resp. pädagogisch-psychologischen Diagnose von Lernbehinderung rückte in den letzten Jahren vor allem der HAWIK in den Mittelpunkt des Interesses (vgl. Zimmermann et al. 1971). Dabei sind grundsätzlich zwei Zielaspekte zu unterscheiden: das *Klassifikationsproblem* (Frage nach der "Sonderschulbedürftigkeit" im Sinne der Lernbehindertenschule) und die *sonderpädagogische Indikation* (Strukturanalyse intellektueller Leistungsvariablen Lernbehinderter zwecks Ableitung sonderpädagogischer bzw. spezieller unterrichtlicher Maßnahmen und didaktischer Konzeptionen).

Die Feststellung der Lernbehinderung ist ein *Klassifikationsproblem* und impliziert prinzipiell ähnliche Schwierigkeiten wie die bei der Schuleignungsermittlung für die Sekundarstufe I oder die bei der Studieneignungsermittlung auf der Sekundarstufe II sich ergebenden. Traditionell wurde – und wird auch heute noch vielfach – das Modell der Grenzwertmethode, zumeist mit einem Cut-off-Score von HAWIK-IQ = 90, verwandt; wegen der früher bereits beschriebenen starken Gruppenüberlappung, hier der Hauptschüler- und der Lernbehindertengruppe, im HAWIK oder einem ähnlichen Intelligenztest stellt dieser Ansatz ein höchst unbefriedigendes Verfahren zur Lb-Erfassung dar. Geht man mit Kemmler (1967) davon aus, daß das "Versagen in der Volksschule" *multifaktoriell* bedingt ist, dann bietet sich logischerweise ein multivariates Modell zur Erfassung von Lernbehinderungen (unterschiedlicher Ätiologie und Erscheinungsformen) an. Analog zu unserem AUKL-Ansatz bei der Schuleignungsdifferenzierung von Gymnasiasten, Real- und Hauptschülern (vgl. Allinger & Heller 1972) müßte man Techniken zur Differenzierung von Sonderschülern, z.B. Lernbehinderten und Nicht-Lernbehinderten, Hauptschülern und Lernbehinderten etc., entwickeln. Erste Ansätze dieser Art im Kontext sonderpädagogischer Fragestellungen finden sich bei Amelang & Zimmermann (1966), Lüer (1967), Kornmann (1971) u.a. Allerdings warnen Zimmermann et al. (1971, S. 57 f.) "vor einer voreiligen Anwendung dieser Methode im sog. 'Hilfsschulüberweisungsverfahren' ... wegen der sich ergebenden gesellschafts- und schulpolitischen Implikationen". Somit wird man sich unter Bezug auf den HAWIK in erster Linie auf die faktorenanalytischen Befunde zu diesem Verfahren (in der Anwendung bei lernbehinderten Sonderschülern) stützen müssen: I. *Verbalfaktor* (Ladungen in WT, AV, AW, BE, GF und BO), II. *Handlungsfaktor* (Ladungen in FL, MT, BE, BO und ZS), III. *Schulleistungsfaktor* bzw. "Schulbildung", "Kurzzeitgedächtnis" etc., also wichtige Bestandteile der *Lernfähigkeit* (Ladungen in RD, AW, ZS, AV und BO). Die Faktorenanalysen von Amelang & Zimmermann (1968), Klauer (1969) und – neuerdings – Schmalohr (1971 a) sind in

ihren Befunden nahezu identisch. Und "Kemmler konnte zeigen, daß die auf diesen Faktoren ladenden Intelligenzmaße signifikante Unterschiede zwischen erfolgreichen Schülern und Schulversagern erbrachten. Daher ist auch zu erwarten, daß die Untertests des HAWIK wesentliche Faktoren des Schulerfolgs erfassen. Allerdings dürfen die Resultate der einzelnen Untertests nicht . . . einfach zu einem Gesamtwert aufaddiert werden. Vielmehr sollten die einzelnen Skalen des HAWIK entsprechend ihrer Bedeutung für die Vorhersage der Sonderschulbedürftigkeit gewichtet werden, d.h. Untertests, die sich besonders gut zur Unterscheidung von Sonderschülern und Normalschülern eignen, müßten stärker in den Endwert eingehen als Untertests, die nur wenig zur Unterscheidung beitragen; Untertests, die überhaupt keinen Beitrag zur Differenzierung leisten, müßten völlig unberücksichtigt bleiben. Bei einer derartigen Gewichtung sind auch die Interkorrelationen der Untertests mit in Rechnung zu stellen. Die so gewichteten und aufaddierten Werte gestatten die bestmögliche Klassifikation, d.h. die Fehlentscheidungen werden auf ein Minimum reduziert" (Zimmermann et al. 1971, S. 56 f.). In der neuesten Fassung der HAWIK-Zusatzauswertung von Schmalohr (1971 a) wurden diese Forderungen weitgehend berücksichtigt. Siehe auch das Klassifikationsschema in Graf. 9, das einen doppelten Positionsvergleich des Pb betr. Gesamt-IQ gestattet.

Grafik 9

Summenprozentverteilung des HAWIK-Gesamt-IQ bei Grundschulern und Sonderschulanwärtern – Klassifikationsschema nach Schmalohr (1971 a) (Mit freundl. Genehmigung des Herausgebers)

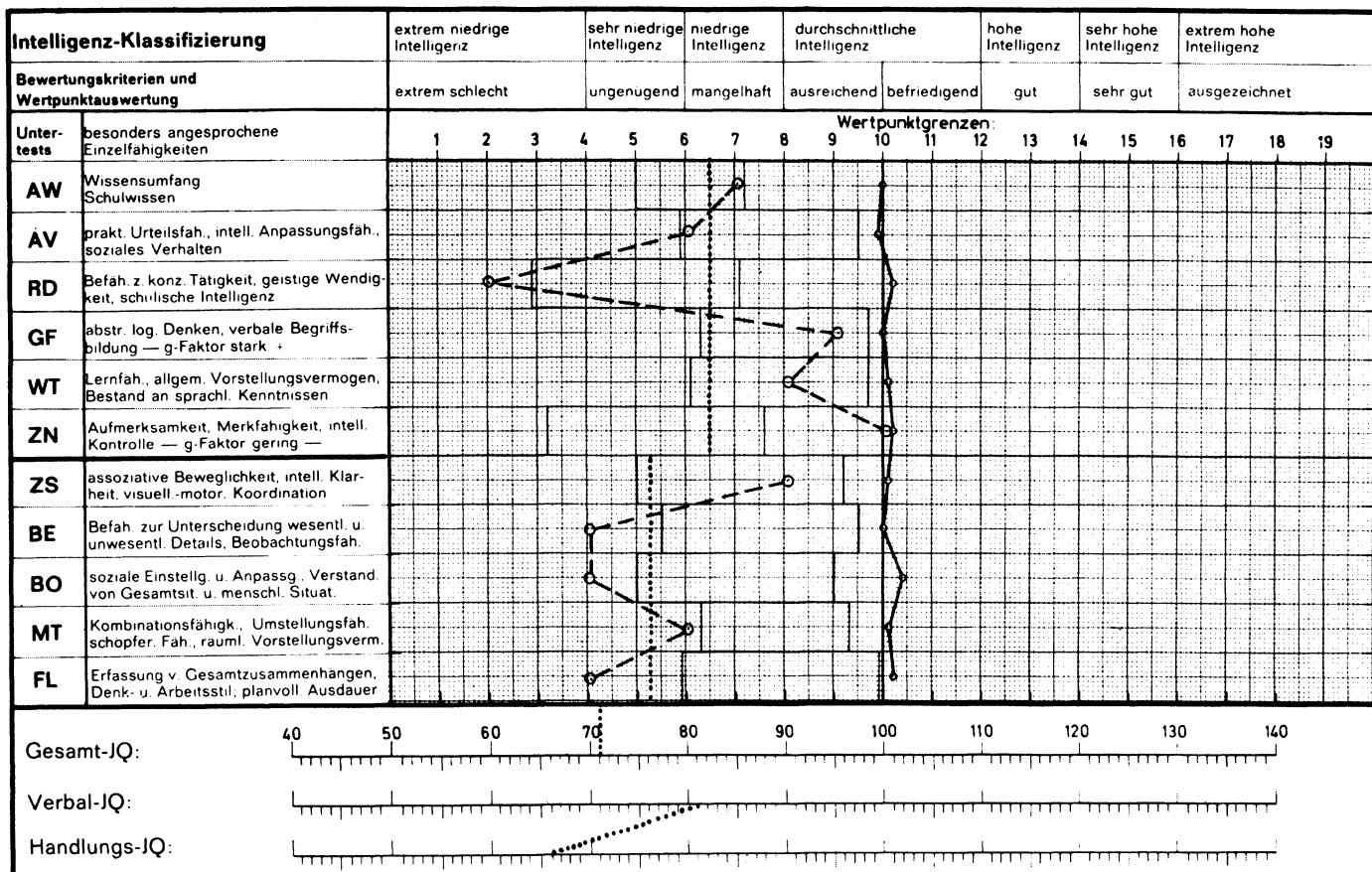


Ein anderer Modus der HAWIK-Auswertung wird durch die Methode der *Profilanalyse* repräsentiert. Für die Lernbehindertendiagnostik im deutschsprachigen Raum liegen hierzu zwei Ansätze vor: 1) die Profilanalyse, die sich auf die *Differenzen im Subtestprofil* des HAWIK stützt (so Schmalohr 1963) und die dem Konzept der sog. kritischen Differenzen folgt (s.S. 157ff. oben). Dabei werden die beobachteten Leistungsunterschiede eines Pb in bezug auf die einzelnen Subtests im HAWIK oder einem vergleichbaren Verfahren auf Zufälligkeit versus Überzufälligkeit hin kontrolliert; signifikante Subtestdifferenzen werden dann im Sinne entsprechender Merkmalsausprägungen interpretiert. Analog bestimmte Schmalohr die kritische Differenz in bezug auf Verbal- und Handlungs-IQ. Schließlich werden bei dem Schmalohrschen Verfahren die "Differenzen vom Profildurchschnitt angegeben: "Dabei handelt es sich um Differenzen, die bei den untersuchten Hilfsschulkindern um eine Standardabweichung (s) positiv und negativ vom Durchschnitt abweichen, d.h. den Bereich der mittleren 68% der Fälle überschreiten" (loc. cit., S. 18).

2) die Profilanalyse, die sich auf den *Vergleich von Individual- und Gruppenprofil(en)* stützt. Diesem Vorgehen liegt die Annahme zugrunde, daß für wohldefinierte Personengruppen, z.B. lernbehinderte Sonderschüler oder auch Gymnasiasten, Real- und Hauptschüler, Durchschnitts- oder Gruppenprofile aufgestellt werden können. "Für alle Unter- tests werden zunächst getrennt Mittelwerte und Verteilungsformen der Punktwerte mit den entsprechenden Kennwerten bestimmt. Mit einem auf diese Weise erstellten normativen Profil können dann in der Praxis die Testleistungen einzelner Versuchspersonen verglichen werden" (Zimmermann et al. 1971, S. 52 f.). In dem von Reinartz & Seifart vorgeschlagenen Verfahren (vgl. Graf. 10) wurde zur Abgrenzung des Mittelbereichs der "wahrscheinliche Fehler" (der vom 1. und 3. Quartil begrenzte mittlere Verteilungsbereich, in dem die mittleren 50% aller Fälle liegen) verwandt. Da jeweils ober- und unterhalb dieser Bereichsgrenzen 25% der Fälle liegen, dürfte hier die Bereichsabgrenzung zu eng gewählt worden sein; heute grenzt man mindestens die mittleren 68% aller Fälle ab (s.S. 171ff. oben). Für Profilinterpretationen der vorgenannten Art sind – wie bereits früher aufgezeigt wurde (s.S. 158) – mindestens folgende Voraussetzungen gefordert: *ausreichende Profilreliabilität* und (damit zusammenhängend) *unterschiedliche Faktorenrepräsentation der Einzel- oder Subtests*. Diesen Anforderungen genügt aber der HAWIK schon in der allgemeinen Anwendungspraxis unzulänglich (vgl. S. 120ff.) und erst recht für die von Schmalohr (ursprünglich) und Reinartz & Seifart vorgeschlagenen spezifischen Auswertungsziele in der Lernbehindertendiagnostik (worauf besonders Zimmermann & Kornmann 1970 hingewiesen haben). Weder die für den HAWIK bei Lernbehinderten gefundenen  $\text{prof}_{\text{Tt}}$ -Werte noch die faktorenanalytischen Ergebnisse zur HAWIK-Leistung Lernbehinderter rechtfertigen die von Reinartz & Seifart vertretene Profilanalyse; zudem müßte man bei dieser Methode eine exaktere Beschreibung der Referenzpopulation vornehmen (als dies durch die genannten Autoren geschieht). Was den Gültigkeitsaspekt des HAWIK-Profiles (unterschiedliche Faktorenrepräsentation der Untertests) Lernbehinderter betrifft, so hat Schmalohr inzwischen (in der 3. Auflage) sein ursprüngliches Auswertungskonzept modifiziert und eine den faktorenanalytischen Befunden (siehe I bis III oben) entsprechende *faktorielle* Profilauswertung der HAWIK-Subtests angeboten. Somit bleibt als letzte Forderung gegenüber einer HAWIK-Subtestprofilauswertung (bei Lernbehinderten) das Postulat einer von *Retest*-Koeffizienten abgeleiteten Bestimmung der  $\text{prof}_{\text{Tt}}$ -Werte (vgl. auch S. 157, Fußn. 51) stehen, da sich auf Halbierungs-Zuverlässigkeitskoeffizienten bezogene Profilreliabilitätswerte im HAWIK (besonders in der Anwendung bei Lernbehinderten) häufig als überschätzt erwiesen (Zimmermann et al. 1971, S. 49 f.).

Wenn wir bei der Erörterung des dritten Anforderungskriteriums an eine HAWIK-Profil- auswertung (Forderung unabhängiger Subtestauswertung durch Verwendung eines einheit- lichen Normmaßstabs, z.B. WP-Normen) bisher stillschweigend davon ausgingen, daß auch in der Lb-diagnostischen Auswertungspraxis die *Wertpunkte* (WP) in den einzelnen HAWIK-Subtests – analog zur Standardisierungsstichprobe Nicht-Lernbehinderter – *normal verteilt* ( $M_{\text{WP}} = 10$  und  $s_{\text{WP}} = 3$ ) seien, dann müssen wir uns jetzt teilweise

# Polygon-Profil zum HAWIK



Grafik 10

Schema zur HAWIK-Profilanalyse lernbehinderter Sonderschüler nach Reinartz & Seifart (1968, 1970<sup>2</sup>)  
(Mit freundl. Genehmigung des C. Marhold Verlags, Berlin, und der Herausgeber)

Tabelle 9

Arithmetische Mittelwerte (M) und Standardabweichungen (s) der Wertpunkt-Verteilungen in den einzelnen HAWIK-Subtests, berechnet aus den WP-Häufigkeiten bei 824 lernbehinderten Sonderschülern – zit. nach Zimmermann et al. (1971, S. 31)

	HAWIK-Subtests										
	AW	AV	RD	GF	WT	ZN	ZS	BE	BO	MT	FL
M	6,24 (6)	7,75 (8)	4,91 (5)	8,13 (8)	7,79 (8)	6,01 (6)	7,46 (8)	7,73 (8)	7,06 (7)	7,61 (8)	7,94 (8)
s	1,88 (2)	2,71 (3)	2,91 (3)	2,66 (3)	2,55 (3)	3,06 (3)	2,92 (3)	2,95 (3)	3,23 (3)	2,40 (2)	3,40 (3)

korrigieren. Nach den von Zimmermann et al. (1971, S. 25 ff.) mitgeteilten Kontrolldaten trifft unsere frühere Annahme in bezug auf Lb nur bei der Hälfte der HAWIK-Subtests (AV, GF, WT, ZS, BE, FL) annähernd zu, was als zusätzliche Kritik an der Profilinterpretation des HAWIK bei Lernbehinderten anzumerken wäre (vgl. Tab. 9).

Somit stellt der HAWIK ein in mancherlei Hinsicht unzulängliches Verfahren dar, wenn es um die Aussonderung von Lernbehinderung geht, wenngleich der Wechsler-Test auch unter diesem Aspekt im Vergleich zu älteren Diagnostikverfahren (z.B. Binet-Test) unverkennbare Vorzüge aufweist. Man wird also noch andere Intelligenztests in Betracht ziehen bzw. auf ihre diagnostische Valenz in der Lernbehindertendiagnostik hin untersuchen müssen. Im folgenden seien als kleiner Beitrag dazu neuere Untersuchungen zum PSB von Horn, AzN 4+ von Hylla & Kraak und CFT von Cattell sowie zu einer Reihe von LU-Variablen referiert. Dabei können wir größtenteils wieder auf eigene Untersuchungsergebnisse, die im Rahmen umfangreicher Begabungsuntersuchungen zur Bildungsberatung 1967/68 angefallen sind, zurückgreifen. Ferner stützen wir uns auf eine sehr interessante Arbeit von Hörr et al. (1972); die Autoren dieser Arbeit haben 1971 eine "annähernd repräsentative" Stichprobe (N = 324) lernbehinderter Sonderschüler der Klassen 7 bis 9 (auf die Klassen 5 und 6 entfielen nur 5 Pbn, so daß wir diese vernachlässigen können) aus dem groß- und mittelstädtischen Raum Mannheim-Heidelberg und der Kleinstadt bzw. Landregion Buchen/Odenw. zusammengestellt. In Tab. 10 werden nun zunächst die Frequenzverteilungen dieser 324 Lb-Pbn neben den Häufigkeitsverteilungen von 212 Lb-Verdachtsfällen aus dem 4. Grundschuljahr eines Großstadtkollektivs – jeweils auf die Subtests im PSB bezogen – aufgeführt.

Die RP-Leistungen der Lb-Sonderschüler (13.-15. Lbj.) sind in allen PSB-Subtests in etwa *normal verteilt*, ausgenommen die Ergebnisse der Subtests 4, 7 und 9, wo schiefe resp. bimodale Verteilungen vorliegen, was aber möglicherweise durch die relativ geringen N-Zahlen und dadurch zu erklärende Zufallseinflüsse bedingt ist. Die Verteilungsergebnisse der jüngeren Viertklässler (Lb-Verdachtsfälle) stellen sich – fast erwartungsgemäß – in den Einzeltests (teilweise) ungünstiger dar, weshalb wir eine nach dem Hornschen Konzept zusammengefaßte Subtestgruppen-Auswertung (1 + 2, 3 + 4, 5 + 6, 7 + 8, 9 + 10) für angemessener erachten. Die entsprechenden Verteilungsergebnisse bezüglich PSB-GL sind Tab. 11 zu entnehmen. Zusammenfassend konstatieren wir, daß der PSB sub specie RP-Verteilung zumindest auf der Oberstufe für Lb-Sonderschüler akzeptabel ist.

Die Frage, ob sich unsere Lb-Verdachtsfälle und die Lb-Sonderschüler in der PSB-Leistung unterscheiden, kann aufgrund der Verteilungsverhältnisse in Tab. 10 u. 11 nicht ohne weiteres entschieden werden. Erst eine Gegenüberstellung der Gruppendurchschnittswerte (M) und Sigmawerte (s), wegen der Pbn-Altersunterschiede jetzt in T-Standard-einheiten ausgedrückt, kann hierüber Aufschluß geben. Die Signifikanzprüfung (F- und t-Test) bezieht sich ausschließlich auf den Gesamtgruppenvergleich "Lb-Sonderschüler" (N = 324) und "LB-Verdachtsfälle" (N = 212); vgl. Tab. 12.



Tabelle 10

Häufigkeitsverteilungen der Rohpunkte in den einzelnen PSB-Subtests bei 212 Lb-Verdachtsfällen des 4. Grundschuljahrs (10. Lbj.) sowie 324 Lb-Sonderschülern der 7. bis 9. Klasse bzw. des 13. bis 15. Lebensjahres; die Verteilungsparameter der Viertkläßkinder entstammen eigenen Untersuchungen, die Statistiken der 13-15jährigen Lb-Sonderschüler wurden der Arbeit von Hörr et al. (1972, S. 110 ff.) entnommen – die Tabellenwerte geben die einzelnen Frequenzen wieder (f)

Altersgruppe (in J.)	N	MP der Klassenintervalle in RP (für jeden Subtest separat aufgeführt)																PSB
		1	4	7	10	13	16	19	22	25	28	31	34	37	40	46	49	1 + 2
10	212	15	34	41	43	28	23	11	9	6	1	0	1					
13	94	1	1	6	9	11	18	17	7	10	3	6	2	1	2	0	0	
14	125	1	3	10	8	14	12	16	13	12	10	12	7	5	1	0	1	
15	105	2	4	8	8	13	10	15	9	12	5	6	6	5	1	0	0	
		0	3	6	9	12	15	18	21	24	27	30	33					3
10	212	2	8	26	55	67	33	12	8	1								
13	94	0	0	2	8	18	16	19	18	11	3	0	0					
14	125	0	0	5	16	21	23	26	21	8	4	1	0					
15	105	0	1	7	7	11	26	19	13	8	10	2	1					
		1	4	7	10	13	16	19	22	25	28	31	34	37	40	43		4
10	212	4	14	29	64	48	32	15	5	1								
13	94	0	0	0	5	22	18	23	21	4	1							
14	125	0	0	4	11	18	37	23	21	9	2							
15	105	0	3	4	8	13	23	17	29	6	2							
10	212	0	1	2	4	21	20	34	33	40	21	17	11	4	3	1		3 + 4
		2	5	8	11	14	17	20	23	26	29	32	35	38	41			5
10	212	4	13	22	44	41	45	23	12	7	1							
13	94	1	1	4	7	11	13	20	12	11	5	6	1	2	0			
14	125	0	1	6	7	15	21	24	21	9	9	10	0	1	1			
15	105	3	6	5	10	9	12	21	12	11	8	1	3	4	0			
		2	5	8	11	14	17	20	23	26	29	32	35	38	41	44		6
10	212	20	50	61	49	21	7	1	1									
13	94	1	5	16	30	19	13	2	4	3	1	0	0	0				
14	125	3	8	25	31	26	13	11	4	4	0	0	0	0				
15	105	4	10	13	24	18	14	12	4	4	0	1	0	1				
10	212	1	2	3	14	20	24	27	35	39	17	15	8	3	2	2		5 + 6
		1	4	7	10	13	16	19	22	25	28	31	34	37	40			7
10	212	79	48	29	19	8	9	9	3	0	3	1						
13	94	10	12	8	5	19	6	5	6	5	5	5	3	4	1			
14	125	14	13	11	18	9	12	5	7	4	9	10	2	7	4			
15	105	15	11	18	10	4	7	7	4	4	3	3	6	11	2			
		1	4	7	10	13	16	19	22	25	28	31	34	37	40	46		8
10	212	23	24	21	27	37	36	24	15	4								
13	94	0	2	2	4	16	22	24	17	5	2	0	0					
14	125	7	4	8	10	22	22	26	13	6	3	4	0					
15	105	6	4	5	8	14	19	17	17	5	8	1	1					
10	212	14	10	17	23	23	27	26	18	14	13	10	7	4	3	3		7 + 8

Alters- gruppe (in J.)	N	MP der Klassenintervalle in RP (für jeden Subtest separat aufgeführt)															PSB
		1	4	7	10	13	16	19	22	25							
																	9
10	212	85	75	28	10	3	1										
13	94	19	12	22	17	17	6	0	1	0							
14	125	15	25	31	17	18	7	8	3	1							
15	105	23	15	20	15	18	8	3	1	0							
		2 <th>7</th> <th>12</th> <th>17</th> <th>22</th> <th>27</th> <th>32</th> <th>37</th> <th>42</th> <th>47</th> <th>52</th> <th>57</th> <th>62</th> <td>10</td>	7	12	17	22	27	32	37	42	47	52	57	62	10		
10	212	26	19	33	32	31	38	19	6	0	2	0	1				
13	94	1	3	1	1	15	20	15	17	9	8	1	3			0	
14	125	4	3	9	10	14	18	20	14	10	12	4	4			3	
15	105	3	2	3	9	12	17	19	19	8	6	4	2			1	
		1 <th>4</th> <th>7</th> <th>10</th> <th>13</th> <th>16</th> <th>19</th> <th>22</th> <th>25</th> <th>28</th> <th>31</th> <th>34</th> <th>37</th> <th>40</th> <th>46</th> <td>9 + 10</td>	4	7	10	13	16	19	22	25	28	31	34	37	40	46	9 + 10
10	212	8	15	9	12	17	26	12	14	23	23	13	19	8	6	5	

Tabelle 11

Rohpunkt-Verteilung der PSB-Gesamtleistung bei 212 Lb-Verdachtsfällen des 4. Grundschuljahrs (10. Lbj.) sowie 324 Lb-Sonderschülern der 7. bis 9. Klasse bzw. des 13. bis 15. Lebensjahrs (siehe auch Tab. 10)

Alters- gruppe (in J.)	MP der Klassenintervalle in RP																	PSB GL
	38	53	68	83	98	113	128	143	158	173	188	203	218	233	248	263	278	
10	2	16	27	50	53	39	11	10	4									f
13	0	0	0	2	0	9	11	18	20	13	10	6	3	1	0	0	1	f
14	0	2	3	2	3	11	9	23	21	20	8	8	8	3	4	0	0	f
15	2	1	3	3	7	9	6	14	13	18	9	7	6	2	2	2	1	f

Tabelle 12

PSB-Leistungen (M- und s-Werte in T) verschiedener Lb-Gruppen und statistische Kontrolle der Gesamtgruppen-Differenzen (N = 324 vs. N = 212)

PSB Variable	Lb-Sonderschüler (13. Lbj. - 15. Lbj.)								Lb-Verdacht		Signif.prüf.	
	N <sub>13</sub> = 94		N <sub>14</sub> = 125		N <sub>15</sub> = 105		N <sub>tot.</sub> = 324		N = 212		F-	t- Test
	M	s	M	s	M	s	M	s	M	s		
1 + 2	33,6	6,0	32,6	6,5	32,0	6,6	32,7	6,4	36,7	8,2	ss	ss
3	39,5	8,0	38,5	8,5	40,0	10,0	39,3	8,8	37,4	7,6	ss	ss
4	39,0	10,0	36,5	11,0	37,0	13,0	37,4	11,4	34,6	7,4	ss	ss
3 + 4	39,3	9,0	37,5	9,7	38,5	11,5	38,3	10,0	34,5	6,6	ss	ns
5	40,2	9,0	38,0	9,0	36,5	10,0	38,2	9,3	41,5	9,2	ns	ss
6	35,0	8,0	33,0	8,5	33,5	10,0	33,7	8,8	40,0	7,2	ss	ss
5 + 6	37,6	8,5	35,5	8,7	35,0	10,0	35,9	9,0	37,7	7,2	ss	s
7	43,4	9,0	42,6	9,0	42,0	10,2	42,6	9,4	41,6	6,7	ss	ns
8	45,5	8,0	42,0	10,5	42,6	9,0	43,5	9,3	43,1	8,5	ns	ns
7 + 8	44,4	8,5	42,3	9,8	42,3	9,6	42,9	9,4	41,0	7,1	ss	ss
9	39,5	8,0	39,5	9,0	37,7	7,0	38,9	8,1	38,8	7,0	ss	ns
10	44,0	11,0	40,3	14,0	41,0	12,0	41,6	12,5	41,4	10,1	ss	ns
9 + 10	41,7	9,5	39,9	11,5	42,1	9,5	41,1	10,3	37,2	8,1	ss	ss
GL	39,3	6,2	37,6	6,9	38,0	7,0	38,2	6,7	33,1	5,5	ss	ss

Demnach weisen die Lb-Verdachtsfälle in der 4. Grundschulklasse durchschnittlich bessere Leistungen (nur) in den *sprachlichen* PSB-Dimensionen (1 + 2, 5, 6, 5 + 6) auf, wohingegen die Lb-Sonderschüler der 7. bis 9. Klasse ihrerseits über deutliche Vorteile in den PSB-Dimensionen "Reasoning" (3, 4), "Techn. Begabung" (7 + 8), "Number" (9 + 10) und dem intellektuellen Gesamtleistungsniveau (GL) verfügen. In den anderen PSB-Variablen treten keine bedeutsamen Differenzen in Erscheinung. Diese Befunde sind in mehrfacher Hinsicht aufschlußreich. Einmal unterstreichen sie aufs neue die starke Betonung *verbaler* Fähigkeiten in der (Grund-)Schule – nur die besseren verbalen Fähigkeiten der Lb-Verdachtsfälle erklären m.E. einigermaßen plausibel, weshalb diese Pbn trotz ihres schlechteren Gesamt-IQ nicht schon längst der Lb-Sonderschule zugewiesen wurden. Andererseits lassen jedoch die besseren Testleistungen der Lb-Sonderschüler im 7. bis 9. Schuljahr in PSB-GL sowie einigen Untertesten (vorab in PSB 3; die anderen Differenzen sind wegen irregulärer Verteilungen im PSB 4 und 9 trotz errechneter Signifikanzen nur unter Vorbehalt interpretierbar) erkennen, daß eine adäquate Sonderbeschulung auch oder gerade bei Lernbehinderungen zu intellektuellen Leistungsverbesserungen führen kann (vgl. auch Klauer 1969). Schließlich fällt die insgesamt größere Heterogenität der Lb-Sonderschulgruppe in bezug auf die PSB-Varianz auf (einzige Ausnahme bildet hier die Variable 1 + 2, wo die größere Leistungsstreuung bei den Lb-Verdachtsfällen zu beobachten ist).

Obwohl der PSB also durchweg für Lb-Pbn, besonders im 4. Schuljahr, zu schwer ist (vgl. Tab. 10 u. 11), ermutigen doch die bisher referierten Ergebnisse zu weiteren Kontrollen des PSB. Dafür sprechen auch die von Hörr et al. erstellten Summen-Prozentkurven, die – abgesehen von der Leistung im PSB 7 – in etwa der oben abgebildeten HAWIK-Ogive (Graf. 9) ähneln. Paradigmatisch veranschaulicht hier Graf. 11 die kumulativen Frequenzen von PSB-GL.

Grafik 11

Kumulative Frequenzen (Summenprozentverteilung der Rohwerte) der PSB-Gesamtleistung bei 324 Lb-Sonderschülern der 7. bis 9. Schuljahrklasse – zit. nach Hörr et al. (1972, S. 31)

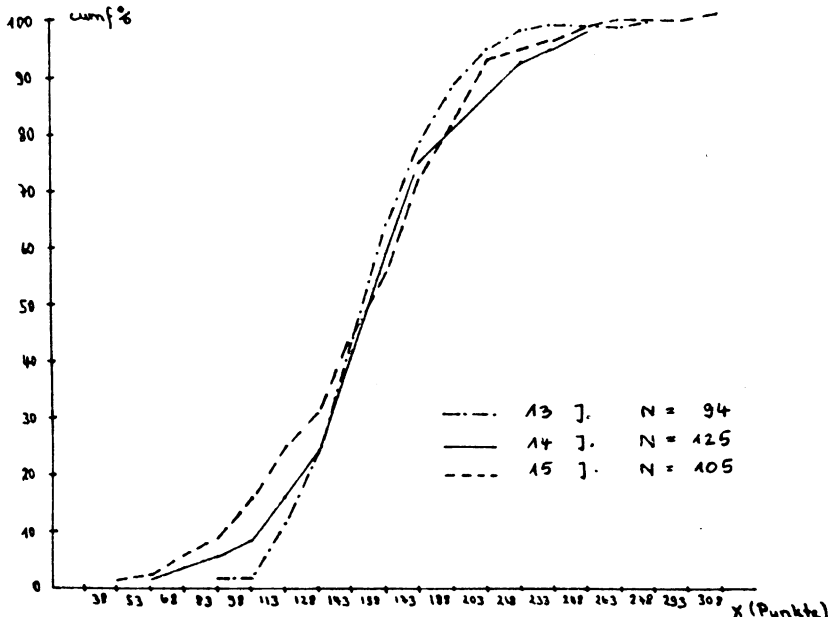


Tabelle 13

Retest- und Paralleltest-Reliabilitätskoeffizienten ( $r_{tt}$  bzw.  $r_p$ ) des PSB in der Anwendung bei 13- bis 15jährigen lernbehinderten Sonderschülern – zit. nach Hörr et al. (1972, S. 70 ff. und S. 84 f.)

PSB	$r_{tt}$ (Retestung nach zwei Wochen)				$r_p$	
	13 J. (N = 33)	14 J. (N = 62)	15 J. (N = 37)	13 - 15 J. Form A      Form B	(N = 324)	
1 + 2	0.89	0.84	0.79	0.82	0.85	0.78
3	0.76	0.70	0.87	0.71	0.83	0.72
4	0.68	0.68	0.85	0.81	0.72	0.70
5	0.74	0.73	0.86	0.73	0.83	0.72
6	0.82	0.69	0.74	0.73	0.75	0.89
7	0.64	0.84	0.84	0.73	0.80	0.75
8	0.39	0.68	0.62	0.52	0.72	0.73
9	0.60	0.68	0.73	0.67	0.64	0.74
10	0.82	0.64	0.66	0.65	0.68	0.63

Für die Lb-diagnostische Praxis von vorrangiger Bedeutung ist die Frage der *Zuverlässigkeit* der PSB-Meßergebnisse; diese interessiert insbesondere auch im Hinblick auf eine mögliche profilanalytische Interpretation der PSB-Resultate Lernbehinderter. In Tabelle 13 sind deshalb die Kontrollergebnisse hierzu mitgeteilt.

Die Reliabilitätskoeffizienten sind zwar fast durchgängig signifikant – die PSB-Formen A und B sind hinreichend parallel, d.h. die oben angeführten Retest-Zuverlässigkeitskoeffizienten der beiden Parallelformen weisen nur zufallsbedingte Differenzen auf –, gleichwohl deuten sie in ihrer Höhe teilweise nur mäßige Stabilität der PSB-Resultate Lernbehinderter an. Erwartungsgemäß ist die Paralleltestzuverlässigkeit ( $r_p$ ) – geringfügig – besser und über die einzelnen Subtests hinweg auch beständiger. Wesentlich günstiger fielen die von Hörr et al. durchgeführten Interkorrelationen der PSB-Subtests aus, die niedrigsten Koeffizienten ergaben sich hierbei für die Lb-Gruppe der 13jährigen, die höchsten für die Lb-Gruppe der 15jährigen (loc. cit., S. 48 ff.). Aus den oben mitgeteilten Reliabilitätskoeffizienten sowie den Ergebnissen der (hier nicht wiedergegebenen) Interkorrelationen erstellen wir nun – analog zu oben S. 158ff. – die Differenzenmatrizes zum PSB bei Lb-Sonderschülern. Dabei sind jeweils die Bezugsdaten ( $r_{tt}$  versus  $r_p$ ) und die *mittleren* Zuverlässigkeitskoeffizienten ( $\bar{r}_{tt}$  versus  $\bar{r}_p$ ) bzw. Interkorrelationsdurchschnittswerte ( $r_{TT}$ ) nebst den wieder errechneten Profilreliabilitäten ( $_{prof}r_{tt}$ ) mitangegeben; siehe Tab. 14 bis 17.

Trotz der eben noch akzeptierbaren Profilreliabilität des PSB (in der Anwendung bei Lernbehinderten) lassen die vergleichsweise hohen Matrix-Werte nur sehr bedingt eine effiziente Profildeutung zu, d.h. nur relativ große Testleistungsdifferenzen können hier intra- versus interindividuell angemessen interpretiert werden, die Erfassung spezifischer Fähigkeiten oder Fähigkeitsstrukturen ist dagegen in den seltensten Fällen möglich (siehe dazu unsere früheren Ausführungen auf S. 159f.). Die geringere Zuverlässigkeit der PSB-Meßergebnisse Lernbehinderter ist in erster Linie ein Effekt überhöhter Aufgabenschwierigkeiten für diese Gruppe: Lb-Sonderschüler lösen bzw. bearbeiten insgesamt weniger PSB-Items als Regelschüler der Primar- oder Sekundarstufe, wodurch notwendigerweise die Testreliabilität in Mitleidenschaft gezogen wird. Über den Weg der Aufgabenmodifikation des PSB für Lernbehinderte, also die Hereinnahme leichterer Testaufgaben und Eliminierung schwieriger, ließe sich ohne weiteres der PSB als brauchbares Testverfahren auch in der Lb-Diagnostik einführen. Die aufgewiesenen Mängel sind zudem kein PSB-Spezifikum, sie treffen mutatis mutandis für alle Intelligenztests, sofern diese an unausgelesenen Populationen standardisiert wurden und nun in der Lb-Diagnostik eingesetzt werden, zu.

Tabelle 14

Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei 13jährigen lernbehinderten Sonderschülern – Sicherung auf dem 5%-Niveau

PSB	1 + 2	3	4	5	6	7	8	9	10	$r_{tt}$	
1 + 2	1,3	2,3	2,6	2,4	2,1	2,7	3,3	2,8	2,1	0.89	$\bar{r}_{tT} = 0.19$
3		1,9	2,9	2,8	2,5	3,0	3,6	3,1	2,5	0.76	
4			2,2	3,0	2,8	3,2	3,8	3,3	2,8	0.68	
5				2,0	2,6	3,1	3,7	3,2	2,6	0.74	$\bar{r}_{tt} = 0.70$
6					1,7	2,9	3,5	3,0	2,4	0.82	
7						2,4	3,9	3,4	2,9	0.64	prof $r_{tt} = 0.64$
8							3,0	3,9	3,5	0.39	
9								2,5	3,0	0.60	
10									1,7	0.82	

Tabelle 15

Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei 14jährigen lernbehinderten Sonderschülern – Sicherung auf dem 5%-Niveau

PSB	1 + 2	3	4	5	6	7	8	9	10	$r_{tt}$	
1 + 2	1,6	2,7	2,7	2,6	2,7	2,2	2,7	2,7	2,8	0.84	$\bar{r}_{tT} = 0.27$
3		2,2	3,1	3,0	3,0	2,7	3,1	3,1	3,2	0.70	
4			2,2	3,0	3,1	2,7	3,1	3,1	3,2	0.68	
5				2,0	3,0	2,6	3,0	3,0	3,1	0.73	$\bar{r}_{tt} = 0.72$
6					2,2	2,7	3,1	3,1	3,2	0.69	
7						1,6	2,7	2,7	2,8	0.84	prof $r_{tt} = 0.64$
8							2,2	3,1	3,2	0.68	
9								2,2	3,2	0.68	
10									2,2	0.64	

Tabelle 16

Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei 15jährigen lernbehinderten Sonderschülern – Sicherung auf dem 5%-Niveau

PSB	1 + 2	3	4	5	6	7	8	9	10	$r_{tt}$	
1 + 2	1,8	2,3	2,4	2,3	2,7	2,4	3,0	2,7	2,9	0.79	$\bar{r}_{tT} = 0.39$
3		1,4	2,1	2,0	2,5	2,1	2,8	2,5	2,7	0.87	
4			1,5	2,1	2,5	2,2	2,9	2,5	2,7	0.85	
5				1,5	2,5	2,1	2,8	2,5	2,7	0.86	$\bar{r}_{tt} = 0.77$
6					2,0	2,5	3,1	2,9	3,0	0.74	
7						1,6	2,9	2,6	2,8	0.84	prof $r_{tt} = 0.64$
8							2,4	3,2	3,3	0.62	
9								2,0	3,1	0.73	
10									2,3	0.66	

Tabelle 17

Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei lernbehinderten Sonderschülern der Oberstufe (Klasse 7-9) – Sicherung auf dem 5%-Niveau

PSB	1 + 2	3	4	5	6	7	8	9	10	$r_p$	
1 + 2	1,8	2,8	2,8	2,8	2,3	2,7	2,7	2,7	3,0	0.78	
3		2,1	3,0	2,9	2,5	2,9	2,9	2,9	3,2	0.72	$\bar{r}_{tT} = 0.29$
4			2,2	3,0	2,5	2,9	3,0	2,9	3,2	0.70	
5				2,1	2,5	2,9	2,9	2,9	3,2	0.72	$\bar{r}_p = 0.74$
6					1,3	2,4	2,4	2,4	2,7	0.89	
7						2,0	2,8	2,8	3,1	0.75	$\text{prof}^{\text{tt}} = 0.63$
8							2,0	2,9	3,1	0.73	
9								2,0	3,1	0.74	
10									2,4	0.63	

Tabelle 18

Testleistungen (M-Werte) der Lb-Verdachtsfälle im AzN 4+ und CFT bei verschiedenen Untersuchungssamples des 4. Grundschuljahrs 1967/68

Stichproben: "Lb-Verdachts- fälle"	Lb-N	AzN 4+ M (Rohpunkte)						M (IQ) GL	CFT M (IQ) GL
		RE	AN	ZR	SE	IV	GL		
Peisert-Reg.* 1 – 19 (Landregionen)	536	1,8	4,1	3,0	1,3	0.9	11,0	64	75
Peisert-Reg.* 1 + 18	49	2,1	4,6	3,0	1,3	0.7	11,0	64	76
Großstadt	20	2,9	7,1	4,5	3,1	2,9	20,4	80	83
Großstadt	217							68	79

\* Zu den sog. Regionen geringerer Bildungsdichte (nach Peisert) vgl. Aurin et al. 1968 sowie Heller 1970 a.

Im Rahmen unserer baden-württembergischen Begabungsuntersuchungen im 4. Grundschuljahr verwendeten wir noch den Frankfurter Übertrittstest AzN 4+ und den Culture Free Test CFT. Für die *Lb-Verdachtsfälle* wurden dabei – je nach Untersuchungssample – oben erwähnten Statistiken ermittelt (vgl. Tab. 18).

Am auffälligsten ist hier wieder die Diskrepanz zwischen den mehr sprachlich betonten Testleistungen im AzN 4+ und der nonverbalen CFT-Leistung aller untersuchten Lb-Verdachtsfälle, und zwar zugunsten der sprachfreien Intelligenzleistung im CFT. Analoge Differenzen wurden auch im Wechsler-Test, dort zwischen Verbal- und Handlungs-IQ zugunsten der Handlungs-Intelligenz, beobachtet (Newman et al. 1955). Hier drängt sich der Verdacht auf, daß sozio-kulturelle Determinanten kovariieren (vgl. auch Estes 1953 u. 1955). Diese Erklärungshypothese gewinnt an Plausibilität, wenn man berücksichtigt, daß sich unsere Lb-Verdachtsfälle hinsichtlich der sozio-ökonomischen Statusvariablen (Schulbildung und Beruf der Kindeseltern) hochsignifikant von den übrigen Grundschulern, also den nicht Lb-Verdächtigen, unterscheiden – mit deutlichen Tendenzen zu den "niedrigeren" Bildungsstufen resp. Berufsklassen (Grundschrift); vgl. dazu ausführlicher Heller 1972 c, S. 23 ff. Hingegen ergaben sich keine bedeutsamen Differenzen bezüglich der Konfessions- und Geschlechtsvariablen. M.a.W.: Phänomene der Lernbehinderung sind im allgemeinen keine geschlechts- oder konfessionsabhängigen Erscheinungen, wohl aber kann Lernbehinderung durch sozio-kulturelle Determinanten verursacht sein.

Wenn wir abschließend die hier referierten Testbefunde zur Lb-Diagnostik noch einmal überblicken, dann müssen wir noch kurz eine bisher kaum beachtete Auffälligkeit ansprechen: die fast durchweg schlechteren Testleistungen im PSB, AzN und CFT im Vergleich zu den eingangs beschriebenen HAWIK-Resultaten. Zwei Erklärungen bieten sich hierfür an. Abgesehen davon, daß die notierten Testwerte teilweise verschiedene Pbn-Gruppen (z.B. Lb-Sonderschüler versus Lb-Verdachtsfälle) repräsentieren, ist einmal wieder der Effekt der Kurzsuljahre bei den 1967/68 untersuchten Pbn in Rechnung zu stellen (s.S. 178), zum andern aber ist zu berücksichtigen, daß AzN, CFT und PSB jeweils im Gruppenverfahren (als sog. Papier-Bleistift-Tests) administriert werden. Kornmann et al. (1972) konnten nun zeigen, daß Lb-Pbn bei Gruppentestung signifikant schlechtere Intelligenztestleistungen erzielen als bei individueller Testadministration (z.B. im HAWIK). Diese Tatbestände gilt es abschließend bei einer kritischen Würdigung der referierten Testbefunde zum HAWIK einerseits und zum PSB, CFT und besonders zum AzN 4+ andererseits zu berücksichtigen, wenngleich die Trends in die eine oder andere Richtung keine Aussagen über bessere versus schlechtere Testvalidität dieser oder jener Testverfahren – erlauben. Lb-diagnostische Gültigkeit von Intelligenztests muß durch andere Kriterien bestimmt werden (siehe dazu u.a. Zimmermann et al. 1971). Immerhin lassen sich in bezug auf unsere via PSB, AzN und CFT ermittelten Lb-Verdachtsfälle einige LU-Außenkriterien als Kontrollmerkmale bzw. Gruppencharakteristika benennen: Die Lb-Anwärtergruppe (N = 212) unterschied sich von allen Nicht-Lb-Anwärtern (N = 3545) durch sehr signifikant schlechtere Schulleistungszensuren in den Hauptfächern (Aufsatz, Rechtschreiben, Rechnen) sowie hochsignifikant schlechtere "Benotungen" hinsichtlich der Arbeitshaltung (Anstrengungsbereitschaft), Konzentrationsfähigkeit, Selbständigkeit und Einordnungsbereitschaft (Sozialverhalten).

## b) Probleme und Ergebnisse der Hörgeschädigten-/Sprachbehindertendiagnostik

Unter "Hörgeschädigte" seien in Anlehnung an die sonderpädagogische Terminologie – wenigstens im Bereich des Hör-Sprachgeschädigtenwesens – sowohl *Gehörlose* (Tauben) und *Resthörige* als auch *Schwerhörige* subsumiert. Dabei nehmen die Resthörigen oder Hörrestler, die in bezug auf ihren Hörverlust und im Hinblick auf die Notwendigkeit sonderpädagogischer (sprachpädagogischer) Maßnahmen weder den Tauben noch den Schwerhörigen eindeutig zugeordnet werden können, eine gewisse Mittelstellung zwischen den Extremvarianten der unter Hörschädigung subsumierten Personengruppen ein, obwohl sie im traditionellen Schulsystem gewöhnlich in der Gehörlosen- oder Taubstummenschule zu finden sind. Freilich sind die Übergänge zwischen den definierten Klassen der Hörschädigung allemal fließend und somit unsere drei Kategorien nur als grobe Ordnungsschemata zu verstehen.

Der Grad der Hörschädigung (vorab im Frequenzbereich der Sprache), allgemein in db-Verlusten<sup>55</sup> formuliert, und die durch die Hörschädigung bedingte Sprachbeeinträchtigung, d.h. Sprachlosigkeit versus Sprachbehinderung bzw. Sprachverzögerung, dienen im allgemeinen als Kriterien der Gruppendiskriminierung. Gelegentliche Vorschläge für die Hereinnahme eines dritten Kriteriums zur Gruppeneinteilung, nämlich der Begabung oder Intelligenz, sind im obigen Differenzierungsmodell nicht berücksichtigt. Eine solche Vernachlässigung mag im Rahmen schulpädagogischer und bildungsorganisatorischer Bemühungen mit Recht bedauert werden, im Hinblick auf unseren spezifischen Untersuchungsgegenstand – die Erfassung des intelligenten Verhaltens Hörgeschädigter – mußte die Variable "Intelligenz" als Gruppierungsfaktor unberücksichtigt bleiben.

<sup>55</sup> Dezibel (Abk. db oder dB) bezeichnet eine (Standard-) Maßeinheit zur Bestimmung der Intensitäts-Unterschieds-Schwelle (Lautstärke) eines akustischen Reizes. Theoretische Voraussetzung hierfür ist die Gültigkeit des Weberschen Gesetzes. Die meisten Audiometer, die zur Erfassung von Hörverlusten heute eingesetzt werden, verwenden die db-Standardskala, um die Gradminderung des (physiologischen) Hörvermögens auszudrücken. Siehe noch Drever u. Fröhlich 1968, S. 56 u. 255.

Schließlich sind wir noch eine Erklärung darüber schuldig, weshalb im folgenden – wie auch schon früher – “Hörgeschädigte” und “Sprachbehinderte” mitunter in eine Gruppe zusammengefaßt werden. Dieses Vorgehen ist gewiß unter *pädagogischen* Gesichtspunkten in mancherlei Hinsicht fragwürdig, und die geschätzten Fachkollegen – sowohl der Hörgeschädigtenpädagogik als auch der Sprachheil- oder Sprachbehindertenpädagogik – mögen hier einmal ihre diesbezüglichen Bedenken zurückstellen. Im Zusammenhang unserer Problemerkörterung geht es ja nicht primär um pädagogische oder schulorganisatorische Fragen, sondern um *intelligenzdiagnostische* Probleme. Diese sind aber sub specie Methode für beide Behinderungsgruppen weithin identisch, weshalb wir auf eine getrennte Behandlung Hörgeschädigter (gehörloser bzw. resthöriger versus schwerhöriger Schüler) und Sprachbehinderter verzichteten. Ohnehin ist mit jeder gravierenden Hörbeeinträchtigung im Kindesalter eine mehr oder weniger starke Sprachbehinderung verbunden. Für die Nicht-Sonderpädagogen unter den Lesern sei hier nur soviel noch angemerkt, daß sich in den Sprachbehindertenschulen (zumeist nach *Sprachheilschulen* bezeichnet) überwiegend Stammer, Stotterer, Polterer etc., aber auch Aphasiker, Dysgrammatiker, Legastheniker und andere sprachauffällige Kinder und Jugendliche befinden (vgl. dazu u.a. Heller 1972 a).

Für die Personengruppe der Hör-/Sprachgeschädigten fehlt es wie kaum für eine andere sonderpädagogisch zu betreuende Gruppe bislang an geeigneten Intelligenzmeßverfahren. Dank der selbstlosen Unterstützung zahlreicher Fachpädagogen<sup>56</sup> sowie aufgrund eigener empirischer Untersuchungen konnte ich nun in den letzten Jahren zahlreiche Testdaten sammeln. Ferner standen einige – z.T. von mir selbst betreute – neuere Examensarbeiten (vgl. u.a. Bangen 1965, Batt 1967, Brennecke 1962, Brunner et al. 1966, Büscher 1964, Eckrich 1970, Holfeld 1968, Kopfmann 1967, Lang 1967, Martin 1961, Schollmeyer 1970) zur Verfügung, die durch andere Arbeiten, besonders von Bäumler et al. 1968, Evans 1966, Gebhardt 1968, Seifert 1960, wertvolle informatorische Ergänzung erfuhren. Darüber wird im einzelnen noch zu berichten sein. Eine wesentliche Bereicherung intelligenzdiagnostischer Möglichkeiten der hier angesprochenen Zielgruppe dürften die im Anhang dieses Buches beigefügten *Hörgeschädigten-Normen* bedeuten, die ohne die erwählte vielfältige Kooperation durch Lehrer-Studenten und Sonderpädagogen nicht in dieser Form zustande gekommen wären. Trotzdem konnten wichtige Anwendungsbereiche nicht oder nur am Rande thematisiert werden, so etwa intelligenzdiagnostische Probleme in der Früherziehung Hör-Sprachgeschädigter. Hierzu verweisen wir insbesondere auf die Veröffentlichungen von Klinghammer (1968) und Neumann (1968). Von den älteren Publikationen seien schließlich noch die einschlägigen Arbeiten von Herderschée (1920), Lindner (1925), Krenberger (1927), Ziehen (1933), Höfler (1934) sowie Zeckel und v.d. Kolk (1939) erwähnt, die jedoch praktisch nur noch historische Bedeutung haben.

Ergänzend zu unseren früheren Ausführungen auf Seite 125ff. sei hier zunächst eine faktorenanalytische Untersuchung zum SON in der Anwendung bei taubstummen Oberstufenschülern von Bäumler et al. (1968) referiert. Die Autoren untersuchten 66 taube/resthörige (Kriterium: 75 db) Schüler – 42 männliche und 24 weibliche Pbn – im Alter von 13 bis 15 Jahren; die Schüler besuchten zum Zeitpunkt der Testerhebung die Oberstufe (6.-9. Klasse) an den Taubstummenschulen in Würzburg, Nürnberg und Essen. Es wurden ausschließlich geburtstaupe bzw. vor der Spracherlernung (also vor dem 2. Lbj.)

<sup>56</sup> An den Testerhebungen beteiligten sich folgende Gehörlosen- und Schwerhörigen-schulen in Heidelberg, Heidelberg-Neckargemünd, Heilbronn, Nürtingen, Waldkirch und Waldshut (alle Baden-Württemberg), Homberg (Hessen), Wuppertal (Nordrhein-Westfalen) und Frankenthal (Rheinland-Pfalz). Den Direktoren der genannten Schulen sowie allen beteiligten Lehrkräften und Schülern sei auch an dieser Stelle für ihre Mitarbeit gedankt. Für die freundliche Überlassung eigener Arbeits- und Informationsunterlagen schulden wir Frau S. Pitsch (Nürtingen) sowie den Herren Dr. H. Ding (Heidelberg), W. Eckrich (Waldkirch), jetzt Freiburg-Stegen), H. Gebhardt (Wuppertal) und G. Lang (Heilbronn) besonderen Dank.



erlaubte, d.h. volltaube oder resthörige Kinder, die keine zusätzlichen Zerebralschäden aufwiesen, berücksichtigt. Neben dem SON wurde noch der HAWIK-Subtest "Rechnerische Denken" (RD) eingesetzt, der "zur Prüfung einer speziellen Art schulrelevanter Denkaufgaben, die im Snijders-Test nicht erfaßt werden", diente. Die Original-Aufgaben des Subtests RD wurden für das Sprachverständnis Hörgeschädigter von TOL Weck (Würzburg) textlich umgearbeitet (Bäumler et al. 1968, S. 89 f.):

1. Ich schneide einen Apfel einmal durch. Wieviele Teile bekomme ich dann?
2. Fritz hat vier Pfennige. Die Mutter schenkt dem Fritz zwei Pfennige. Wieviel Pfennige hat er dann?
3. Otto hat acht Kugeln. Er kauft noch sechs dazu. Wieviele Kugeln hat er dann?
4. Ein Mann hat 12 Zeitungen. Er verkauft fünf Zeitungen. Wieviele Zeitungen hat er noch?
5. Eine Zigarette kostet sieben Pfg. Wieviel kosten drei Zigaretten?
6. Ein Milchmann hat 25 Flaschen Milch. Er verkauft 11 Flaschen. Wieviele Flaschen hat er noch?
7. Vier Buben haben zusammen 72 Pfg. Sie teilen das Geld. Wieviel bekommt jeder Bub?
8. Ein Bub hat 36 Mark verdient. Er hat für einen Tag vier Mark bekommen. Wieviele Tage hat er gearbeitet?
9. Ich kaufe drei Dutzend (= 36) Druckknöpfe. Ein Dutzend (= 12) kostet 30 Pfg. Ich gebe eine Mark. Wieviel bekomme ich heraus?
10. Drei Schreibfedern kosten fünf Pfg. Wieviel kosten 24 Schreibfedern?
11. 36 ist  $2/3$  von ? oder  $2/3$  von ? ist 36?
12. Ein Taxi kostet für den ersten Viertelkilometer 20 Pfg. Für jeden weiteren Viertelkilometer kostet es fünf Pfg. Ein Mann fährt zwei Kilometer mit dem Taxi. Wieviel muß er bezahlen?
13. Herr Müller und Herr Schulze beginnen ein Kartenspiel. Jeder hat 27 Mark. Sie versprechen: Wer ein Spiel verliert, der muß  $1/3$  von seinem Geld bezahlen. Herr Müller gewinnt drei Spiele. Herr Schulze verliert 3mal. Wieviel Geld hat Herr Schulze noch?

Die einzelnen Aufgaben waren auf Kärtchen gedruckt, die dem Pb vorgelegt wurden; sobald die Aufgabenstellung verstanden war, mußten die Kärtchen abgegeben werden.

Für die Leistung in RD ergaben sich ein Gruppenmittelwert von  $M_{RP} = 9,1$  und eine Abweichung von  $s_{RP} = 2,32$  (jeweils in Rohpunkt-Einheiten ausgedrückt). Für die Leistung im SON wurden  $M_{IQ} = 100,4$  und  $s_{IQ} = 14,34$  (jeweils in Abweichungs-IQ formuliert), also praktisch identische Werte zur SON-Standardisierungsstichprobe, ermittelt. Die Rohwerte aller 9 Teilstests wurden anschließend interkorreliert und faktorisiert (Hauptachsenmethode, Varimax-Rotation, Kommunalitätsschätzung: Quadrat d. multipl. Korr.koeff.); auf der Basis von Standardwerten ergaben sich praktisch identische Resultate. Die Korrelationsmatrix ist in Tab. 19 wiedergegeben.

Tabelle 19

Interkorrelationen zum SON und RD (HAWIK) bei 66 tauben/resthörigen Schülern im Alter von 13 bis 15 Jahren – zit. nach Bäumler et al. (1968, S. 90)

	Mosaik	Zeichn.	Komb.	Ergänz.	Analog.	Sort.	Gedä.	Knox	RD
Mosaik	—	.727	.465	.401	.609	.236	.286	.461	.503
Zeichnen		—	.454	.436	.644	.255	.269	.443	.364
Kombination			—	.505	.539	.386	.406	.266	.448
Ergänzung				—	.407	.341	.435	.428	.267
Analogien					—	.401	.369	.310	.559
Sortieren						—	.216	.313	.567
Gedächtnis							—	.425	.326
Knox Würfel								—	.238
RD (HAWIK)									—

Mit Ausnahme der Korrelationen zwischen Sortieren/Gedächtnis für Karten und Knox Würfel/RD (HAWIK) sind alle Korrelationen auf dem 1%-Niveau signifikant.

Aus der – hier nicht angeführten – Faktorenmatrix ergaben sich drei sinnvoll interpretierbare Faktoren. “Mehr als drei bedeutsame Faktoren sind auch nach den Eigenwerten nicht zu erwarten. Die Eigenwerte der 5 extrahierten Faktoren sind  $F_1 = 4,075$ ;  $F_2 = 0,828$ ;  $F_3 = 0,740$ ;  $F_4 = 0,460$ ;  $F_5 = 0,356$ ” (loc. cit., S. 91). Die interpretierten Faktoren sind:

Faktor I	= “Erfassen von geometrischen, figuralen (optisch dargebotenen) Regelmäßigkeiten” mit den höchsten Ladungen in
	Mosaiktest .822
	Figurenergänzung .834
	Figurenanalogien .620
	Kombination .377
Faktor II	= “Erkennen von Sinnzusammenhängen auf vorwiegend begrifflich-abstraktem Niveau” mit den höchsten Ladungen in
	RD (HAWIK) .806
	Sortierten .740
	Analogien .449
	Kombination .413
Faktor III	= “Aufmerksames Beobachten und daraus resultierend besseres unmittelbares Behalten von optischem Reizmaterial” mit den höchsten Ladungen in
	Gedächtnis f. Karten .702
	Knox Würfel .656
	Bilderergänzung .629
	Bilderreihen (Komb.) .404

Die von Bäumler et al. bei hörgeschädigten Oberstufenschülern eruierte SON-Faktorenstruktur deckt sich demnach nur teilweise mit der von Snijders & Snijders-Oomen getroffenen Subtestgruppierung. Am ehesten noch lassen sich Faktor I und die SON-Dimension (sensu Snijders-Oomen) “Form”, Faktor II und die SON-Dimension “Abstraktion” sowie Faktor III und die SON-Dimension “Unmittelbares Gedächtnis” zuordnen, während zur zweiten SON-Kategorie sensu Snijders-Oomen (“Anschaulicher Zusammenhang”) von Bäumler et al. kein Faktor ermittelt werden konnte.

Als praktische Konsequenz ihrer faktorenanalytischen Untersuchungsbefunde schlagen Bäumler et al. vor, künftig nur noch folgende *drei Subtestreihen des SON* zu verwenden: *Mosaik-Test*, *Karten-Sortieren* und *Gedächtnis* für Karten. Zunächst eingeschränkt auf die Oberstufe der Gehörlosenschule (und Schwerhörigenschule?) gültig bedeutet diese Reduzierung immerhin eine erhebliche Vereinfachung des SON-Diagnostikums ohne Informationsverluste bzw. intelligenzdiagnostische Einbußen. Ob die vorgeschlagene Rationalisierung des SON auch in der Anwendung auf der Grundschulstufe effizient ist, müßten analoge Faktorenanalysen erst noch ergeben. Bis dahin sollte man den SON bei Vor- und Grundschulkindern im Sinne eines “allgemeinen” Intelligenzmeßverfahrens benutzen (siehe oben S. 125f.).

Zu den am meisten in der Hörgeschädigtendiagnostik eingesetzten Intelligenztests gehört der *PMT von Raven*, obwohl dieser Test eigentlich nicht speziell für diese Zielgruppe geschaffen wurde. Als nonverbaler Test eignet er sich jedoch augenscheinlich recht gut für Intelligenzuntersuchungen Hör-/Sprachgeschädigter. Abgesehen von einigen älteren Informationen (z.B. Seifert 1960, Heller 1967 b) fehlen aber bislang genauere Angaben zur Kontrolle des PMT in der (ausgedehnten) Anwendung bei Hörgeschädigten sowie gruppenspezifische Testnormen, deren Notwendigkeit sich bereits in einer an englischen Taubstummen von Evans 1966 durchgeführten PMT-Untersuchung herausstellte. Auf der Basis von insgesamt 1260 PMT-Protokollen deutscher Hörgeschädigter im Alter von 5;0 bis 19+ J. können wir nunmehr unsere seinerzeit (1967) veröffentlichten PMT-Normen erheblich verbessert anbieten. Die Ergebnisse sind in Tab. I a (Altersnormen von 5 bis

19+ J.), Tab I b (Schulnormen für Taube), Tab. I c (Schulnormen für Schwerhörige), Tab. 1 d (Hörgeschädigten-Normen für Real- und Wirtschaftsschüler) sowie Tab. I e (verschiedene Berufsschulnormen für Hörgeschädigte) im Anhang dieses Buches zusammengefaßt. Die Kontrolle der sozio-ökonomischen Statusvariable (Vaterberuf) in unseren Normierungsstichproben erbrachte folgende Verteilungsergebnisse für das Gehörlosensample bzw. Schwerhörigensample: 4,0% bzw. 6,5% Akademiker; 4,4% bzw. 6,8% Beamte und Angestellte der gehobenen Laufbahn, Journalisten, Offiziere etc.; 13,1% bzw. 12,7% Angestellte und Beamte, Gewerbetreibende u.ä.; 8,2% bzw. 8,7% Selbständige Handwerker, Handwerksmeister u.ä.; 7,0% bzw. 4,0% Bauern; 36,9% bzw. 37,8% Industriefacharbeiter, einfache Angestellte und Beamte usw.; 7,8% bzw. 6,2% Angelernte, nicht-technische Berufe; 11,4% bzw. 13,0% Ungelernte Arbeiter; 7,2% bzw. 4,3% Sonstige (Rentner, Hausfrauen usw.). Unsere Eichstichproben dürften somit hinsichtlich der sozio-ökonomischen Statusvariablen als einigermaßen repräsentativ gelten.

Ergänzend zu den bereits oben (s.S. 134f.) mitgeteilten Kontrollen der PMT-Gütekriterien geben wir hier eine Gesamtübersicht der Untersuchungen zur *Validität* des PMT in der Anwendung bei hörgeschädigten Kindern und Jugendlichen. Aus Rationalisierungsgründen geben wir die entsprechenden Befunde zum SON und HAWIK-Handlungsteil sogleich mit an; vgl. Tab. 20.

Aus den Tabelleninformationen lassen sich folgende Gültigkeitsaspekte des PMT, SON und HAWIK-Handlungsteils – jeweils in der Anwendung bei hörgeschädigten Kindern und Jugendlichen – ableiten. PMT (HD), d.h. die auf unseren (Heidelberger) Hörgeschädigten-Normen basierenden PMT-Resultate tauber und resthöriger bzw. schwerhöriger Pbn, und SON korrelieren je nach Untersuchungssample zwischen 0.53 und 0.90 miteinander, während PMT (R), d.h. die auf den Raven-Normen für Normalhörende basierenden PMT-Resultate, und SON zwischen 0.50 und 0.80 miteinander korreliert sind. Entsprechend ergaben sich Zusammenhänge zwischen PMT (HD) und HAWIK-Handlungsteil zwischen 0.53 und 0.87, in der Relation PMT (R) und HAWIK-Handlungsteil jedoch nur zwischen 0.27 und 0.66. SON und HAWIK-Handlungsteil sind dagegen zwischen 0.60 und 0.80 miteinander korreliert. Zu den Schulleistungen weisen PMT und SON etwa gleich hohe Zusammenhangswerte auf, wenngleich die PMT-Leistung am besten mit der Rechenzensur (0.60 bis 0.68) bzw. dem Gesamt-Notendurchschnitt (0.63 bis 0.91) übereinstimmt und die SON-Leistung am besten mit der Deutschzensur (0.74). Das Ergebnis im HAWIK-Handlungsteil steht demgegenüber mit  $r = 0.61$  (Deutschzensur) und  $r = 0.36$  (Rechenzensur) nur in einem mäßigen Zusammenhang mit der Schulleistung.

Die dargestellten Validitätsbefunde liegen also im üblichen Rahmen, teilweise (besonders beim PMT und SON) sogar erheblich über den sonst gefundenen Validitätskoeffizienten einschlägiger Intelligenztests. Insgesamt korrelieren die PMT (HD)-Werte besser mit den jeweiligen Außenkriterien als die PMT (R)-Werte, was nicht nur eine Erhöhung der PMT-Validität – bei entsprechender Verwendung hörgeschädigtenspezifischer Normen – bedeutet, sondern zugleich – indirekt – die Berechtigung bzw. Forderung populationspezifischer Testnormen überhaupt unterstreicht.

Zu den PMT (HD)-Normen selbst wäre noch anzumerken, daß sich nach unseren Berechnungen keine signifikanten Differenzen bezüglich der PMT-Leistung in den Alterspopulationen der Gehörlosen (Tauben) und Resthörigen/Schwerhörigen nachweisen ließen, wohl aber in der Schuljahrgruppierung zwischen den einzelnen Klassen und Schülergruppen. Deshalb wurden hierfür nach Hörschädigungsgrad getrennte PMT-Normen berechnet (Tab. I b u. I c), während die Altersnormen für die Gesamtgruppe der Hörgeschädigten gelten (Tab. I a). Ebenso beziehen sich die PMT-Normen in Tab I d und Tab I e, vorab aus Gründen der Stichprobengröße bzw. Rpräsentanz, auf die Gesamtgruppe.

Die bereits angeführten Argumente bezüglich einer möglichen Erhöhung der testdiagnostischen Valenz durch Verwendung populationspezifischer Auswertungsnormen waren schließlich ausschlaggebend dafür, weitere Hörgeschädigten-Testnormen zu erstellen.

Tabelle 20

Ergebnisse zur Gültigkeitsbestimmung des PMT, SON und HAWIK-Handlungsteils in der Anwendung bei Hörgeschädigten

Stichprobe	N	Beziehung	r	Signifikanz	Quelle
7;6-10;5 T	23	PMT(HD) – SON	0.53	ss	Lang 1967
10;6-13;5 T	18	PMT(HD) – SON	0.58	s	"
13;6-16;5 T	17	PMT(HD) – SON	0.89	ss	"
7;6-16;5 T	58	PMT(HD) – SON	0.67	ss	"
Taube Schüler	97	PMT(HD) – SON	0.66	ss	Heller
Taube Schüler	93	PMT(HD) – SON	0.90	ss	"
Taube Schüler	169	PMT(HD) – SON	0.57	ss	"
Schwerhörige Schüler	27	PMT(HD) – SON	0.76	ss	"
Schwerhörige Schüler	76	PMT(R) – SON	0.50	ss	"
7;6-10;5 T	23	PMT(HD) – HAWIK-Handl.	0.61	ss	Kopfmann 1967
10;6-13;5 T	18	PMT(HD) – HAWIK-Handl.	0.53	s	"
13;6-16;5 T	17	PMT(HD) – HAWIK-Handl.	0.80	ss	"
7;6-16;5 T	58	PMT(HD) – HAWIK-Handl.	0.60	ss	"
Taube Schüler	22	PMT(HD) – HAWIK-Handl.	0.87	ss	Heller
Taube Schüler	21	PMT(HD) – HAWIK-Handl.	0.70	ss	"
Schwerhörige Schüler	112	PMT(R) – HAWIK-Handl.	0.66	ss	"
5;6-10;5 T	45	PMT(R) – WISC-Handl.	0.27	(s)	Evans 1966
10;6-15;5 T	55	PMT(R) – WISC-Handl.	0.44	ss	"
Taube Schüler	77	PMT(R) – SON	0.80	ss	Bangen 1965
Taube Schüler	133	PMT(R) – LU	0.57	ss	"
9;6-10;5 T	18	PMT(R) – SL (tot.)	0.67	ss	"
9;6-10;5 T	13	PMT(R) – SL (tot.)	0.91	ss	"
13;6-14;5 T	21	PMT(R) – SL (tot.)	0.63	ss	"
13;6-14;5 T	16	PMT(R) – SL (tot.)	0.67	ss	"
Schwerhörige Schüler	70	SON – HAWIK-Handl.	0.80	ss	Brennecke 1962
Taube Schüler	23	SON – HAWIK-Handl.	0.60	ss	Heller
Schwerhörige Schüler	42	SON – HAWIK-Handl.	0.68	ss	"
Taube Schüler	58	SON – HAWIK-Handl.	0.78	ss	"
Taube Schüler	69	PMT(HD) – SL (Spr. + Abs.)	0.54	ss	"
Taube Schüler	69	PMT(HD) – SL (Absehen)	0.40	ss	"
Taube Schüler	69	PMT(HD) – SL (Sprechen)	0.41	ss	"
Taube Schüler	68	PMT(HD) – SL (tot.)	0.64	ss	"
Taube Schüler	125	PMT(HD) – SL (Rechnen)	0.68	ss	"
Taube Schüler	67	PMT(HD) – LU (Fl. + Arb.)	0.29	s	"
Taube Schüler	26	SON – SL (Spr. + Abs.)	0.74	ss	"
Taube Schüler	52	SON – SL (Rechnen)	0.59	ss	"
Taube Schüler	134	PMT(R) – SL (Rechnen)	0.60	ss	Büscher 1964
Taube Schüler	134	PMT(R) – SL (Deutsch)	0.58	ss	"
Taube Schüler	16	HAWIK-Handl. – SL (Spr. + Abs.)	0.61	s	Heller
Taube Schüler	16	HAWIK-Handl. – SL (Rechnen)	0.36	ns	"
Taube Schüler	67	SL (tot.) – LU (Fl. + Arb.)	0.60	ss	"

T = Taube; PMT(HD) = Heidelberger PMT-Normen (vgl. Tab. I); PMT(R) = PMT-Normen nach Raven (für Hörende); SL = Schulleistungszensur(en); SL(Spr.) = Schulleistungszensur in Sprechen; SL(Abs.) = Schulleistungszensur im Absehen; SL(Spr. + Abs.) = Schulleistungszensur in Deutsch; SL(tot.) = Schulleistungszensuren insgesamt (Zensurendurchschnitt); LU = Lehrerurteil; LU (Fl. + Arb.) = Lehrerurteil bezüglich Fleiß und Arbeitshaltung etc.

Die entsprechenden Ergebnisse zum *Benton-Test*, die aus einer Untersuchung von Batt (1967) resultieren (siehe oben S. 140f.), und zum *HAWIK-Handlungsteil*, die auf der Grundlage von über 200 HAWIK-Protokollen tauber/resthöriger Pbn von uns berechnet wurden, finden sich in Tab. II a u. II b sowie Tab. III im Anhang. Im Hinblick auf die relativ geringen N-Zahlen der Normstichprobe(n) und damit verbundene mögliche Repräsentanzmängel möchten wir diese Normdaten vorerst mehr im Sinne einer Arbeitsgrundlage verstanden wissen. Diese Einschränkung gilt praktisch auch für alle folgenden Normentabellen (Tab. IV bis Tab. VII). In der diagnostischen Praxis wird man deshalb vielfach nicht darauf verzichten können (und wollen), jeweils für die betr. Schul- und/oder Klassengruppe *eigene Normdaten* zu erstellen. Dabei ist eine Normierung praktisch nur auf Ordinalskalenniveau (Rangskala bzw. %-Rangskala = PR) möglich. Die am Schluß unseres Tabellen-Anhangs beigefügte Tab. IX zur Umwandlung von Rangplätzen in T-Standardwerte wird – bei sachgerechter Benutzung – den sonst üblichen Arbeitsaufwand auf ein Minimum reduzieren und somit die Gewinnung von Kleingruppen-Normen wesentlich erleichtern. Ihre Verfügbarkeit stellt gerade in bestimmten klinischen und sonderpädagogischen Anwendungsbereichen für die (Intelligenz-)Diagnostik eine unerläßliche Voraussetzung dar, nicht zuletzt unter dem Aspekt der Gültigkeit bzw. Effizienz diagnostischer Aussagen. Und praktisch die gleichen Prinzipien gelten für eine objektive, zuverlässige und gültige *Leistungsbeurteilung* in der Schule, deren Problemkomplex jedoch hier nicht thematisiert ist.

Trotz der aufgezeigten Testgütekriterien und inzwischen vielfach bestätigten Brauchbarkeit haftet aber allen bisher behandelten Verfahren zur Messung der Intelligenz Hörgeschädigter ein erheblicher Mangel an: als sog. nonverbale Tests erfassen sie spezifische *verbale* Fähigkeiten nur über deren Beteiligung an der "allgemeinen" Intelligenz oder/und an bestimmten Primärfähigkeiten, z.B. *Reasoning*komplex. Speziell für Hörgeschädigte fehlt bislang ein *verbaler* Intelligenztest. Deshalb ist eine erste Erprobung einschlägiger Tests, hier der WST 5-6 und WST 7-8 sowie des FTU 4-6, bei Hör-/Sprachgeschädigten von besonderem Interesse. Darüber soll im folgenden berichtet werden.

Fast gleichzeitig haben Eckrich (1970) und der Verfasser erste umfangreichere WST-Erhebungen an Gehörlosen- und Schwerhörigenschulen der BRD durchgeführt. Daß sie dabei immer wieder auf erhebliche Bedenken seitens der Sonderpädagogen und Schulen stießen, war nur allzu verständlich und soll hier am Rande doch vermerkt werden. Aber uns ging es ja nicht um die Erfassung schulischer Lernleistungen oder gar Lehrerkontrollen, wie des öfteren vermutet wurde, sondern um die Erprobung eines Wortschatztests, also eines verbalen Intelligenztests oder allenfalls eines kombinierten Lern-Intelligenztests, wobei *einzelne* Klassen- oder Schulgruppenresultate überhaupt nicht interessierten und auch keine entsprechenden Leistungsanalysen vorgenommen wurden.

Erwartungsgemäß erwies sich der *Frankfurter Wortschatztest* WST 5-6 (mit den Parallelformen I a und I b) bzw. WST 7-8 (mit den Parallelformen II a und II b) in der vorliegenden Form für hörgeschädigte Schüler der vorgesehenen Klassenstufen als zu schwierig. Die von Eckrich errechneten Schwierigkeitsindizes belegen dies insgesamt recht eindrucksvoll<sup>57</sup>. Andererseits konnten die WST-Ergebnisse wichtige empirische Hinweise für eine Testmodifikation resp. Neukonstruktion eines verbalen Intelligenztests für Hörgeschädigte liefern, und das war sogar der Hauptzweck unserer Untersuchung.

Wenn wir trotzdem in Tab. IV WST-Schulprozenträge für Hörgeschädigte anbieten, so geschieht dies vorab zu dem Zweck, dem Gehörlosen- bzw. Schwerhörigenpädagogen grobe Orientierungsmaßstäbe für die WST-Leistung Hörgeschädigter und damit eine erste

<sup>57</sup> Die von Eckrich (loc.cit.) monierten "Schönheitsfehler" im Aufbau des WST stören hier zunächst weniger; so ergab eine Analyse der Wortarten bei allen vier Testformen (Ia bzw. Ib und IIa bzw. IIb) signifikante Verteilungsunregelmäßigkeiten der Wortkategorienwahl (Substantive, Verben, Adjektive, Adverben etc.) bezüglich der angebotenen Antwortauswahlmöglichkeiten, nicht aber bezüglich der sog. Schlüsselwörter (Stimulusitems).

Arbeitsgrundlage zur *verbalen Intelligenz*-Testung Hörsprachgeschädigter zu liefern. Die in Tab. IVa niedergelegten Normdaten machen deutlich, daß auf der Hauptschulstufe der *Gehörlosenschule* unter gewissen Vorbehalten praktisch nur der WST 5-6 (Form I) als diagnostisches Provisorium in Betracht kommt, während sich der Einsatz des WST 7-8 (Form II) durchweg verbietet, da maximal kaum 1/3 der (60) Testitems hier gelöst wird. Aus denselben Gründen sollte bei *schwerhörigen* Hauptschülern der WST 5-6 durchgängig bevorzugt werden. Die vorgenannten Normdaten gelten nur im Hinblick auf die Verwendung des WST als Powertest, d.h. ohne zeitliche Vorgabenbegrenzung.

Lediglich bei den hörgeschädigten *Real-* und *Wirtschaftsschülern* ergeben sich – bei geringfügig erhöhter Bearbeitungszeit von regulär 15 Min. auf hier 20 Min. – annähernd normale Verteilungen der WST-Leistung, in bezug auf Testform II (WST 7-8) allerdings erst ab der 8. Klassenstufe. Die PR-Angaben bei den Wirtschaftsschulklassen sind wegen minimaler Stichproben-N hier mehr von theoretischer als praktisch-diagnostischer Bedeutung.

Die Anwendung des *FTU 4-6* bei schwerhörigen Schülern wurde erstmals von Schollmeyer (1970) erprobt. "Die Gruppe der Versuchspersonen, mit denen der FTU 4-6 in den Jahren 1968 und 1969 durchgeführt wurde, stellt eine Auslese der an der Schwerhörigenschule in Waldkirch für den Besuch der Realschule befähigten Schüler dar. Die Kriterien für die Auslese waren die bisherigen Schulleistungen und ein Gutachten des jeweiligen Klassenlehrers, in dem besonders der sprachliche Entwicklungsstand eines jeden Schülers eingehend dargestellt war. Art und Grad der Hörschädigung und der mit dem HAWIK-Handlungsteil ermittelte IQ wurden für die Auslese nicht berücksichtigt" (loc.cit., S. 17). Insgesamt wurden 34 (20 männliche und 14 weibliche) Pbn mit einem mittleren Hörverlust zwischen 43 und 75 dB untersucht, wobei sich folgende Ergebnisse abzeichnen.

1. Die FTU-Leistungen der untersuchten schwerhörigen Schüler unterscheiden sich nicht signifikant von denen der normalhörenden Realschüler, sofern – und diese Einschränkung muß beachtet werden – die um 1 Jahr zurückverschobenen Testnormen, also die FTU-Normen der um 1 Jahr jüngeren Normalhörenden, als Vergleichsmaßstab dienen.
2. Schüler, die seit der 1. Klasse (1. Schuljahr) die Schwerhörigenschule besuchen, zeigen sich gegenüber den erst später (aus der Grundschule bzw. Hauptschule, Realschule, Gymnasium) in die Schwerhörigenschule eingewiesenen Schülern im FTU leistungsüberlegen.
3. Unter Bezugnahme auf die FTU-Ergebnisse folgert Schollmeyer, daß die sprachlichen Voraussetzungen für die erfolgreiche Teilnahme am Englischunterricht bei schwerhörigen Schülern nach etwa 5 Schulbesuchsjahren gegeben seien. Eo ipso ist diese Aussage – einschränkend – nur für jene schwerhörigen Schüler gültig, die auch sonst den (allgemeinen) intellektuellen und Lernleistungsvoraussetzungen, soweit diese Kriterien der Stichprobenbildung darstellten (siehe Tab. 21), genügen.
4. Insgesamt betrachtet bestehen nur mäßige Zusammenhänge zwischen der FTU-Leistung und dem HAWIK-Handlungs-IQ, die praktisch nur in der Beziehung FTU-Subtest 1 und HAWIK-Handlungsteil bedeutsam werden. Relevante Zusammenhänge zwischen dem Grad der Hörschädigung (mittlerem dB-Verlust) und FTU-Leistung manifestieren sich praktisch nur im FTU-Subtest 5 und – in geringerem Maße – im FTU-Subtest 1 sowie in der FTU-Gesamtleistung. Streng genommen gilt diese Aussage wiederum nur unter Berücksichtigung der von Schollmeyer genau beschriebenen Versuchs- bzw. Administrationsbedingungen.
5. Eine relativ enge Beziehung offenbart die Korrelation von FTU-Gesamtleistung und Englischzensur (0.6), wohingegen der HAWIK-Handlungs-IQ auf die Englischleistung praktisch keinen Einfluß hat (vgl. Tab. 21).

Aufgrund einer Analyse der Schwerhörigen-Ergebnisse in den einzelnen FTU-Subtests vermutet Schollmeyer, "daß Minderleistungen in den Bereichen der Syntax und des Lesens vorliegen. Eine Anhebung der Leistungsfähigkeit in diesen Bereichen durch unterrichtliche Maßnahmen scheint möglich zu sein. Die Minderleistung im Bereich des auditiven Erfassens

Tabelle 21

Interkorrelationen des FTU 4-6 mit verschiedenen Außenkriterien bei drei schwerhörigen Untersuchungssamples (Gesamt-N = 34) – eigene Berechnungen nach den Untersuchungsdaten von Schollmeyer (1970)

Beziehung		Stichpr.-N.	Rangreihen- Korr.koeff.	Signifikanz
FTU-GL (T)	– $\phi$ dB-Verlust	33	0.41	s
FTU-Subtest 1(RP)	– $\phi$ db-Verlust	31	0.43	s
FTU-Subtest 2(RP)	– $\phi$ db-Verlust	31	–0.03	ns
FTU-Subtest 3(RP)	– $\phi$ db-Verlust	31	0.10	ns
FTU-Subtest 4(RP)	– $\phi$ dB-Verlust	31	–0.14	ns
FTU-Subtest 5(RP)	– $\phi$ dB-Verlust	31	0.49	ss
FTU-GL (T)	– HAWIK-Handl. (IQ)	34	0.17	ns
FTU-Subtest 1(RP)	– HAWIK-Handl. (IQ)	33	0.73	ss
FTU-Subtest 2(RP)	– HAWIK-Handl. (IQ)	33	–0.26	ns
FTU-Subtest 3(RP)	– HAWIK-Handl. (IQ)	33	–0.25	ns
FTU-Subtest 4(RP)	– HAWIK-Handl. (IQ)	33	0.11	ns
FTU-Subtest 5(RP)	– HAWIK-Handl. (IQ)	33	–0.07	ns
Engl.-Zensur (LU)	– HAWIK-Handl. (IQ)	19	0.00	–
Engl.-Zensur (LU)	– FTU-GL (T)	19	0.61	ss

sinnloser Silben hängt unmittelbar mit der Art und dem Grad der Hörschädigung zusammen und ist auf die Unangemessenheit der Aufgaben des fünften Testteils zurückzuführen“ (loc.cit., S. 42). Daraus ergibt sich die Notwendigkeit, den FTU 4-6 für die Anwendung bei schwerhörigen Pbn umzuarbeiten und neu zu standardisieren. In der Arbeit von Schollmeyer finden sich dazu bereits konkrete Vorschläge, die bei einer entsprechenden Modifikation des FTU 4-6 für Hörgeschädigte berücksichtigt werden sollten.

Aus den schon mehrfach angesprochenen Gründen einer umfassenden Intelligenzdiagnose schien es angebracht, bei Begabungsuntersuchungen Hörgeschädigter relevante *Persönlichkeits-* und *Interessendimensionen* mitzuerfassen. Zu diesem Zwecke erprobten wir den Persönlichkeits-Interessen-Test (PIT) von Mittenecker & Toman und den Berufs-Interessen-Test (BIT) von Irle bei hörgeschädigten Oberstufenschülern. Die Ergebnisse werden an anderer Stelle noch eingehend zu besprechen sein. Hier seien wiederum – analog zu den BIT-Resultaten der H-, R- und G-Schülergruppen Normalhörender (vgl. S. 170) – die Rangfolgen der bei den einzelnen Schulgruppen bevorzugten Interessendimensionen im BIT vergleichend gegenübergestellt.

- Hörgeschädigte *Hauptschüler* der Kl. 8-10 (N = 57):  
 GH, KB, TH, EH, LG, VB, SE, LF, TN
- Hörgeschädigte *Realschüler* der Kl. 7-10 (N = 48):  
 GH, KB, EH, VB, LG, SE, TN, TH, LF
- Hörgeschädigte *Wirtschaftsschüler* der Kl. 10-12 (N = 28):  
 VB, EH, KB, LG, GH, TN, TH, SE, LF
- Hörgeschädigte *Real- und Wirtschaftsschüler* (N = 76):  
 KB, VB, EH, GH, LG, TN, SE, TH, LF

Im Gegensatz zu den BIT-Resultaten der Normalhörenden sind hier bei den Hörgeschädigten kaum interschulische Differenzen bezüglich der Rangfolge der einzelnen Interessendimensionen erkennbar. KB, EH, GH und VB werden von allen Schulgruppen hier deutlich bevorzugt, während sich die Präferenzen bei den oben beschriebenen Schulpopulationen Hörender kontrastreicher darstellten. Allerdings muß bei der Interpretation der insgesamt weniger profilierten Interessenstrukturen Hörgeschädigter in Rechnung gestellt werden, daß der BIT doch für manche Pbn sprachliche Schwierigkeiten bereitete, so daß die Zuverlässigkeit und Validität des BIT in der Anwendung bei Hörgeschädigten nicht ohne

weiteres gewährleistet sind. Sofern genügende Sprachkenntnisse beim Pb vorausgesetzt oder tutorische Unterstützung des Sprachverständnisses einzelner Testitems gesichert werden können, vermittelt der BIT ab der 9. bzw. 10. Klassenstufe auch bei Hörgeschädigten fallweise wertvolle Informationen über berufsbezogene Interessenlagen.

Neben Voraussetzungen sprachlicher Kompetenz und im engeren Sinne kognitiver Faktoren (der Intelligenz) werden immer auch sozio-kulturelle Milieufaktoren an der Intelligenztestleistung Hörgeschädigter beteiligt sein. Nach den Untersuchungsergebnissen von Büscher (1964) und Holfeld (1968) kann jedoch angenommen werden, daß die Bedeutung *familiärer* Erziehungs- und Bildungseinflüsse auf die intellektuelle Entwicklung Hörgeschädigter anstelle schulischer Erziehungs- und Bildungseinwirkung vergleichsweise in den Hintergrund tritt. Grad und Ausmaß vorliegender Sinnesschädigung bestimmen hier vielfach die reduzierten Möglichkeiten familiärer Bildungsbemühungen versus die Notwendigkeit *sonderpädagogischer* Ansätze. Dies wird im nächsten Abschnitt bei der Behandlung blindenpsychologischer Probleme noch einmal deutlich in Erscheinung treten. Unter dem Methodenaspekt kehren sich freilich die Intelligenzdiagnostischen Probleme beider Behinderungsgruppen (der Hör- versus Sehgeschädigten) geradezu um: Während hier die angemessene Erfassung *sprachlicher* Fähigkeiten einige Mühe bereitet, avanciert in der Blindendiagnostik die *nonverbale* (haptische) Intelligenzmessung zum Methodenproblem Nr. 1.

### Exkurs: Das Legasthenie-Syndrom

Zu den häufigsten sprachlichen Störungen in der Schule gehört die *Lese-Rechtschreib-Schwäche* oder sog. Legasthenie. Sie ist gekennzeichnet durch die "Inkongruenz von allgemeiner Begabungshöhe und der Fähigkeit, das Lesen und Schreiben in der von der Schule dafür eingeräumten Zeit und mit dem vorgesehenen Maß an Training zu erlernen. Wir haben hier also den eigentümlichen Fall vor uns, daß auf eine drastische Weise die gemessene Intelligenz auf der einen Seite und der Schulerfolg in den wesentlichen Bereichen des Lesens und Schreibens auf der anderen Seite nicht kovariieren." (Schubenz & Buchwald 1964, S. 155). Unabhängig von bislang ungeklärten theoretischen Streitigkeiten über Ursachen und Bedingungen dieser Verhaltensauffälligkeit tritt hier die *diagnostische* Problematik in den Vordergrund, d.h. die Frage nach den Unterscheidungsmerkmalen auffälligen versus unauffälligen Verhaltens im oben definierten Sinne. Klasen (1971), die ausführlich über die diagnostischen Probleme informiert, stellte eine Reihe einschlägiger Testuntersuchungen zusammen. Dabei ergab sich im HAWIK bzw. HAWIE, der wohl am häufigsten als Intelligenztest hierbei Verwendung findet, folgendes Bild (loc.cit., S. 213):

Schiffman & Clemmens (N = 240)	M (Verbal-IQ) = 91,0; M (Handl.-IQ) = 101,0
Graham (N = 65)	M " = 88,4; M " = 101,7
Altus (N = 25)	M " = 97,8; M " = 100,4
Kallos et al. (N = 37)	M " = 95,9; M " = 103,0
Graham (N = 31)	M " = 98,9; M " = 101,7
Robeck (N = 36)	M " = 106,7; M " = 111,7
Klasen (N = 488)	M " = 104,3; M " = 105,5

Mit Ausnahme der Klasenschen Untersuchung tritt demnach eine mehr oder weniger deutliche Diskrepanz von Verbal- und Handlungs-IQ zugunsten des letzteren bei den HAWIK-Resultaten legasthenischer Schüler in Erscheinung. Allerdings sind die aufgezeigten Differenzen in den seltensten Fällen statistisch bedeutsam. In dem Untersuchungsmaterial von Klasen tritt nur in 41,2% der getesteten Legastheniker eine signifikante Differenz (in 22,3% der Fälle zugunsten eines höheren Verbal-IQ und in 18,9% der Fälle zugunsten eines höheren Handlungs-IQ) zutage. Die von der Autorin für die beiden Gruppen berechneten mittleren Subtest-Standardwerte (WP) – Profil I = Verbal-IQ signifikant höher, Profil II = Handlungs-IQ signifikant höher – lauten (loc.cit., S. 216):

	AW	AV	RD	GF	WT	ZN	BE	BO	MT	FL	ZS
Profil I	12,2	12,8	10,7	12,9	12,1	10,5	10,1	9,6	10,1	9,5	9,2
Profil II	8,7	9,6	7,9	9,7	9,4	8,8	11,9	11,4	11,7	11,8	11,0



Ohne Berücksichtigung der Diskrepanzen zwischen Verbal- und Handlungs-IQ ergaben sich die nachstehend in Tab. 22 wiedergegebenen HAWIK-Subtestprofile, ausgedrückt in WP-Einheiten, für 466 Legastheniker.

Tabelle 22

Subtestprofile (Mittelwerte in WP) im HAWIK bei 466 Legasthenikern, differenziert nach Geschlecht und Schulstufe – zit. nach Klasen (1971, S. 190)

HAWIK-Subtest	Gesamt-Stichprobe N = 466	Knaben N = 333	Mädchen N = 133	Primarschüler N = 352	Sekundarschüler N = 114
AW	10,50	10,76	9,84	10,59	10,21
AV	10,96	11,15	10,49	10,71	11,74
RD	9,45	9,68	8,89	9,52	9,23
GF	11,41	11,56	11,24	11,35	11,60
WT	10,91	11,23	10,83	11,37	9,48
ZN	9,42	9,38	9,53	9,43	9,37
BE	10,83	11,12	10,10	10,88	10,67
BO	10,76	10,87	10,47	10,78	10,69
MT	10,76	11,10	9,90	10,69	10,95
FL	10,65	10,80	10,26	10,59	10,83
ZS	9,84	9,54	10,59	9,62	10,53

Die Legastheniker offenbaren vor allem im RD und ZN deutliche Leistungsschwächen im Wechsler-Test, während sie in den Subtests AV, GF, BO, MT und FL eher über dem Durchschnitt liegen. Entsprechende Profilvergleiche zwischen Individualprofil (legasthenieverdächtigem Schülerprofil) und einem der oben angegebenen Gruppenprofile – mit oder ohne Berücksichtigung der V-IQ/H-IQ-Diskrepanz – sind freilich in bezug auf den HAWIK nur bedingt aussagekräftig, wie die früheren Auseinandersetzungen mit dem Wechsler-Test zur Genüge darlegten. Für die praktische Diagnostik der Legasthenie bedeutet diese Einschränkung, daß man sich niemals auf die HAWIK-Symptome allein stützen darf und kann. Für eine intensivere Beschäftigung mit den hier angeschnittenen Problemen empfehlen wir die Lektüre von Klasen (1971), wo auch weitere Literaturhinweise zu finden sind.

### c) Probleme und Ergebnisse der Sehgeschädigtendiagnostik\*

Im Lichte der modernen Begabungsforschung enthält unsere Thematik zwei zentrale Fragestellungen: a) die Frage nach den unmittelbaren Auswirkungen von Blindheit resp. Sehbehinderung auf die intellektuelle Ausstattung und Entwicklung im Kindes- und Jugendalter und b) die Frage nach möglichen (aufweisbaren) Zusammenhängen zwischen visueller Deprivation und intellektueller Verhaltensleistung. Beide Frageaspekte – der “biologische” und der “kognitive” – werden im Kontext der Mehrfachbehinderung relevant.

\* Der folgende Abschnitt ist unter dem Titel “Sehgeschädigung und Intelligenz” bereits in den Beiträgen zur Mehrfachbehindertenpädagogik (Hrsg. N. Hartmann) im G. Schindele Verlag, Karlsruhe-Neuburgweiler 1972, S. 185-201, erschienen. Der Abdruck erfolgt mit freundlicher Genehmigung des Schindele Verlages.

Eine Beantwortung dieser Fragen stößt zunächst auf erhebliche *methodische* Schwierigkeiten, die bereits Karl Bühler 1913 (auf dem 14. Blindenlehrerkongreß in Düsseldorf/Düren) in den Grundzügen skizzierte, deren Lösungsansätze noch 60 Jahre danach sehr oft nicht befriedigten. Zwar forderte derselbe Autor 1921 auf der 1. Fortbildungswoche für österreichische Blindenpädagogen sehr dezidiert die Entwicklung spezifischer Blindentests, wobei er auf die (für Sehende) einschlägigen Arbeiten von Binet und Simon sowie der deutschen Forscher Bobertag, Ziehen und Stern hinwies, doch ist das Anliegen "blind-seinsgemäßer" Diagnostika – von wenigen Ausnahmen abgesehen – bis in die jüngste Vergangenheit hinein Desiderat geblieben. Zudem enthält die Schaffung spezieller Tests für Blinde und (hochgradig) Sehbehinderte ihre besondere Problematik hinsichtlich einiger Grundannahmen der Blindenpsychologie, vorab unter wahrnehmungs- und persönlichkeitspsychologischen versus sozialpsychologischen Fragestellungen (vgl. J. Schmidt 1958, 1960).

Die meisten bisherigen Intelligenzuntersuchungen Sehgeschädigter beschränkten sich auf die Verwendung der bekannten Binet- oder Wechslerkalen, wozu in der Regel allerdings keine speziellen Eichunterlagen zur Verfügung standen. Standardisierungsansätze, die die Kriterien der modernen Teststatistik hinreichend berücksichtigen, finden sich erst in allerjüngster Zeit (siehe u.a. H. Horn 1970). Dies bedeutet, daß den hier referierten Intelligenzbefunden Blinder und Sehbehinderter vielfach nur *vorläufige* Aussagekraft zukommt.

Bereits vor dem Zweiten Weltkrieg wurden in den USA Intelligenzuntersuchungen bei Blinden durchgeführt. Größere Bedeutung erlangten die Forschungen S.P. Hayes', der mit seinen "Interim Hayes-Binet Tests for the Blind" (IHB), einer für Blinde adaptierten Stanford-Binet-Revision von 1937, in den Jahren 1941-47 insgesamt 482 blinde und hochgradig sehbehinderte Schüler untersuchte. Gleichzeitig wandte er bei 324 blinden Schülern den Verbalteil der Wechsler Adult Intelligence Scale (WAIS) an und berechnete Korrelationen um 0.9 zwischen IHB und WAIS (Verbal). Nachdem Hayes noch 1941 einen zweijährigen Rückstand der intellektuellen Entwicklung und allgemeinen Schulleistungsfähigkeit blinder Kinder konstatiert hatte, wobei "blind children tend to test lower than the seeing in all Binet items which one would classify under thinking or reasoning" (S. 280 f.), spezifizierte er seine Aussagen rund 20 Jahre später folgendermaßen: "In both comparisons (WAIS und IHB; d. Verf.) it will be observed that the blind children show smaller per cents in the middle ranges of intelligence, but higher per cents than the seeing at both the superior and the inferior levels" (1962, S. 145 f.). Heute findet zunehmend die Wechsler Intelligence Scale for Children (WISC) bei Intelligenzuntersuchungen Sehgeschädigter Verwendung. Über weitere blindendiagnostische Untersuchungsverfahren informieren Dautermann et al. (1967). Siehe auch S. 114ff., 123f., 141f. u. 146f. in diesem Buch.

Im deutschsprachigen Raum sind nennenswerte Intelligenzuntersuchungen Blinder und Sehschwacher, von sporadischen Ansätzen der Vorkriegszeit (etwa den Kombinations- und Definitionsaufgaben, die Voß 1922 an der Provinzialblindenanstalt Kiel einführte – und als Mittel der Polemik gegen die Testpsychologie seiner Zeit verwandte) abgesehen, vergleichsweise jungen Datums. So berichtete Mansholt 1961 über seine bei 110 blinden und hochgradig sehbehinderten Schülern im Alter von 11 bis 15 Jahren durchgeführten Begabungsuntersuchungen und stellte eine Retardierung der blinden Schüler "bei der Entfaltung ihrer geistigen Begabung, vor allem in den Bereichen des Sachverhaltsmerkennens und der Beziehungserfassung" fest. Im gleichen Berichtsjahr publizierte Strehle (1961 a, 1961 b) seinen, auf einer Dissertation des englischen Arztes Drummond aus dem Jahre 1918 basierenden, Intelligenz-Punkt-Test für Blinde (IPT). Strehle selbst hat bei der Übertragung der Binet-Vorlage einzelne Items der 22 Aufgabenreihen im Hinblick auf deutschsprachige Verhältnisse modifiziert. Im Anschluß daran unternahm Rexhausen (1964) eine Eichung des IPT, wozu ihm die entsprechenden Protokolldaten von 155 vollblinden und 170 teilblinden (hochgradig sehbehinderten?) Schülern als Berechnungsgrundlage dienten. Auf Einzelergebnisse dieser beiden Arbeiten werden wir später noch zurückkommen.

Eine der interessantesten Studien sehbehindertenpsychologischer Provenienz stellt zweifellos Klauers HAWIK-Untersuchung an der Duisburger Sonderschule für sehbehinderte Kinder (N = 67) Anfang der 60er Jahre dar. Der Autor verfolgte damit einen zweifachen Zweck: Einmal sollte die Frage geprüft werden, "ob und in welchem Ausmaß sich eine Leistungs-minderung des Gesichtssinnes auf die intellektuelle Entwicklung auswirke", zum andern stand das Problem der Anwendbarkeit des HAWIK bei sehschwachen Schülern, etwa die (hypothetische) Frage nach einem typischen HAWIK-Profil Sehbehinderter, zur Entscheidung an. Die Intelligenzuntersuchungen Klauers aus den Jahren 1961/62 gewinnen umso größere Bedeutung, als noch 1958 Anastasi den Mangel repräsentativer Testerhebungen bei *sehbehinderten* Kindern (auch in den USA) beklagte. Daß gerade die Repräsentativität der Klauerschen Untersuchungsergebnisse – neben der Gültigkeit einiger Interpretationshypothesen – heute angezweifelt wird, schmälert keineswegs den Wert dieser Untersuchung, zumal erst durch diese Arbeit eine Reihe weiterführender Ansätze ermöglicht und initiiert wurde.

In den theoretischen Vorüberlegungen zu seiner Untersuchung weist Klauer (1962) auf den engen Zusammenhang zwischen Sehen und Denken hin, der sich schon in unserer Alltagssprache ("Einsicht", "Überblick", "durchsichtig", "klar sehen" u.ä.) ankündigt. Ergänzend möchte ich darauf hinweisen, daß auch dem Begriff "Wissen" in seiner etymologischen Wurzel (lat. videre = sehen → einsehen, erkennen → kennen, wissen) diese Bedeutung inhäriert. Unter Bezugnahme auf unseren eingangs erwähnten zweiten (kognitiven) Frageaspekt wäre in diesem Zusammenhang noch ein Hinweis Buytendijks auf die Etymologie des Wortes "Erfahrung" angebracht. Demnach bedeutet "erfahren" einen Weg immer wieder *intensiv* (durch)fahren, wobei nach Graumann drei Intensitätsstufen unterschieden werden müssen: vom bloßen Zur-Kennntnis-Nehmen über das Wissen (mit mehr oder weniger ausgeprägtem Gewißheitsgrad) hin zum wirklichen Wissen i.S. eigentlicher Erkenntnis, indem unsere *Erfahrung* in der Form von gesehen haben alles Wissen fundiert. Der enge Zusammenhang von Erkennen (Denken versus Intelligenz) und Wahrnehmen (Erfahrung) wird somit offenbar. Im Hinblick auf unsere Untersuchungsthematik stellt sich jetzt die konkrete Frage nach den Auswirkungen restringierter Wahrnehmungs- und Erfahrungsmöglichkeiten im Sinne reduzierter Seh- und damit Einsichts-(Wissens-)Funktionen auf die kognitive, i.e.S. intellektuelle (Einsicht → Intellekt) Entwicklung Sehgeschädigter.

Klauers theoretischer Untersuchungsansatz muß allerdings wesentlich eingegrenzt begriffen werden, insofern er sich in seinen Prämissen auf den Zusammenhang zwischen *Formsehen* und Denken beschränkt. "Wenn es einen Zusammenhang zwischen Formsehen und Denken gibt, kann sich die intellektuelle Entwicklung bei Kindern, deren Formsehen und -erkennen in starkem Maße beeinträchtigt ist, nicht unter den gleichen Bedingungen vollziehen wie bei normal sehtüchtigen Kindern, so daß hieraus schließlich eine spezifische Schädigung der Intelligenz resultieren könnte" (a.a.O., S. 571).

Für das Untersuchungsvorhaben Klauers schien der Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK) "besonders geeignet wegen seiner Gliederung in einen Verbalteil, der den Gesichtssinn nicht unmittelbar beansprucht, und einen Handlungsteil, zu dessen Lösung ein Mindestmaß an Sehfähigkeit vorhanden sein muß". Wechsler selbst betonte schon den Wert der Vergleichsmöglichkeit von menschlicher "Fertigkeit im Umgang mit Worten und Symbolen" (einerseits) mit der "Fähigkeit, Gegenstände zu handhaben und visuelle Gestalten zu erfassen" (andererseits) an Hand der beiden HAWIK-Teile (1956, S. 147). Somit waren die Voraussetzungen zur Überprüfung folgender Arbeitshypothesen gegeben (Klauer, loc.cit.):

- "1. Bei sehschwachen Kindern ist der Handlungs-IQ besonders stark beeinträchtigt.
2. Die Höhe des Handlungs-IQ variiert bei sehschwachen Kindern mit der verbliebenen Sehleistung.
3. Sehschwäche ist eine globale Schädigung, die sich auch auf die verbale Intelligenzleistung mindernd auswirkt.
4. Dementsprechend liegt der Gesamt-IQ sehschwacher Kinder im allgemeinen unter der Norm (Sehender)".

Tabelle 23

Mittelwerte (M) und Standardabweichungen (s) des HAWIK-Verbal-IQ, HAWIK-Handlungs-IQ und HAWIK-Gesamt-IQ bei verschiedenen Stichproben bzw. Visusgruppen sehbehinderter und blinder Schüler<sup>58</sup>

Stichprobe	N	Visusgruppe	Verbal-IQ		Handlungs-IQ		Gesamt-IQ	
			M	s	M	s	M	s
Klauer 1962	31	I	= A*		98,4		99,3	
"	18	II	111,7		96,7		105,2	
"	18	III	117,3		86,4		102,9	
"	67	I+II+III (A+B)	108,0	16,2	94,7	16,2	101,8	15,8
B. & B. 1967	10	I*	102,8		95,8		99,4	
"	18	II	107,2		92,0		99,9	
"	45	III	105,0		87,0		95,5	
"	39	I**	111,4		97,3		105,3	
"	31	II	96,7		83,0		87,5	
"	3	III	109,0		78,3		93,7	
"	73	I+II+III* **	105,2	18,2	88,1	18,7	97,3	18,1
B. & B. 1967 (vgl. Graf. 14)	20	I***	105,2		—		—	
"	19	II	104,5		—		—	
"	23	III	105,2		—		—	
"	62	I+II+III	105,0	18,8	—		—	

Anm.: \* = Binokularer Zentralfer্নvisus (I = 5/15 u. besser, II = 5/20-5/25, III = 5/30 u. schlechter).

\*\* = Binokularer Zentralnahvisus (I = Nieden 1-4, II = Nieden 5-8, III = Nieden 9-12).

\*\*\* = "Visus"-Äquivalent der Blindenstichprobe (I = beim HAWIK-Handlungsteil nicht verwertbarer Sehrest, II = Lichtschein, III = Amaurose).

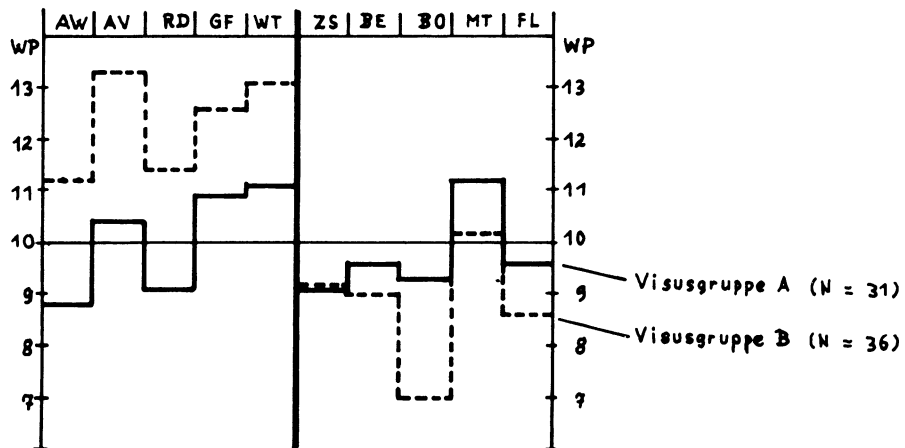
Um das Ergebnis vorwegzubringen: Die ersten beiden Hypothesen konnte Klauer verifizieren; in bezug auf die 3. und 4. Hypothese mußte die Nullhypothese beibehalten werden, d.h. diese Annahmen wurden falsifiziert. Doch weniger an den operationalen Befunden als vielmehr an Klauers Interpretation dieser Befunde entzündete sich jüngst eine lebhaft Diskussion, wie gleich noch zu zeigen sein wird. Zuvor jedoch seien die nicht weniger interessanten Ergebnisse zum HAWIK-Profil Sehbehinderter mitgeteilt (siehe Grafik 12). Zum Vergleich bringen wir anschließend sofort die entsprechenden Resultate von Baitinger & Bernd aus dem Jahre 1967 (siehe Grafik 13).

Den grafischen Darstellungen ist zu entnehmen, daß der HAWIK-IQ in den einzelnen Subtests des Verbalteils fast durchweg über den betreffenden Subtests des Handlungsteils liegt. Für die einzelnen Gesamt- und nach dem Grad der Sehschädigung differenzierten Unterstichproben wurden folgende Mittel- und Sigmawerte ermittelt. (vgl. Tab. 23).

<sup>58</sup> Aus synoptischen Gründen werden die HAWIK-Resultate des Hudelmayerschen Blinden-Samples (s. Grafik 14) mitangegeben.

Grafik 12

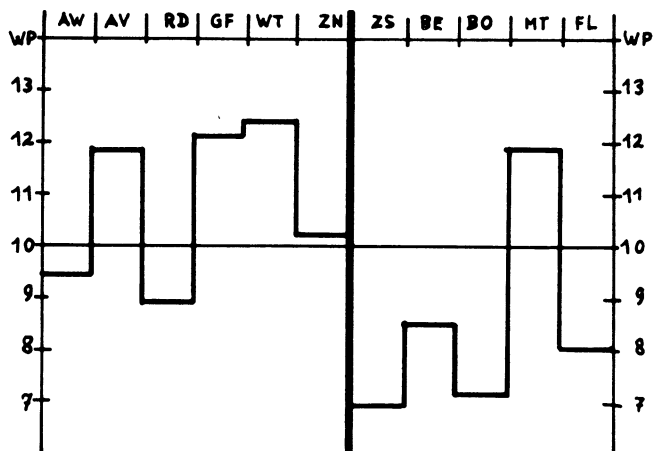
Durchschnittliche HAWIK-Profile sehbehinderter Schüler nach Klauer (1962)



Visusgruppen: A = leichte bis mittelgradige Sehbehinderung, B = hochgradige Sehbehinderung; WP = Wertpunkte ( $M = 10$ ,  $s = 3$ ); Subtests im HAWIK-Verbalteil: AW = Allgemeines Wissen, AV = Allgemeines Verständnis, RD = Rechnerisches Denken, GF = Gemeinsamkeiten finden, WT = Wortschatz-Test, ZN = Zahlennachsprechen; Subtests im HAWIK-Handlungsteil: ZS = Zahlen-Symbol-Test, BE = Bilderergänzen, BO = Bilderordnen, MT = Mosaik-Test, FL = Figurenlegen.

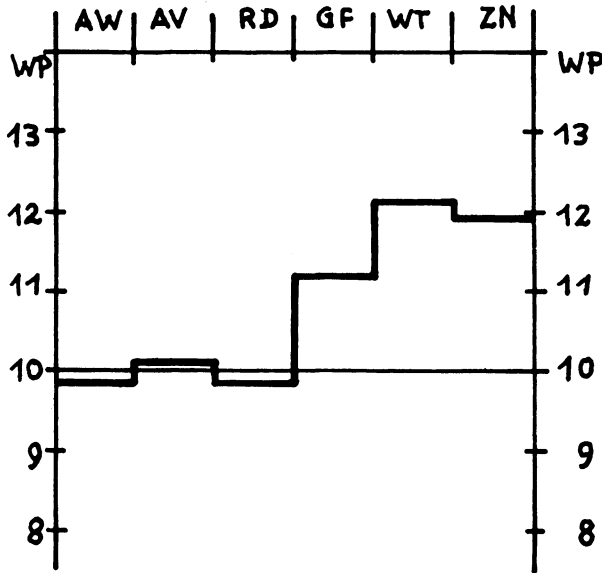
Grafik 13

HAWIK-Gruppenprofil sehbehinderter Schüler ( $N = 73$ ) nach Baitinger & Bernd (1970)



Grafik 14

Gruppenprofil blinder Schüler (N = 62) im HAWIK-Verbalteil nach Hudelmayer (1970, S. 57)



Im Vergleich zu den bei Priester angegebenen Mittel- und Sigmawerten der Standardisierungsstichprobe (Volksschüler)<sup>59</sup> fällt neben den schon erwähnten M-Differenzen die *stärkere Streuung der Sehbehinderten und Blinden* im Verbal-, Handlungs- und Gesamt-IQ auf. Die Abweichungen betragen 3 bis 7 IQ-Punkte und zeigen sich am ausgeprägtesten in den *sprachlichen Subtests*. Ferner ist ein deutliches *Absinken des Handlungs-IQ mit zunehmendem Grad der Sehschädigung* erkennbar, dem (überraschend) in der Klauerschen Untersuchungsstichprobe – nicht ganz so eindeutig in den übrigen Samples – eine *Steigerung des Verbal-IQ bei abnehmender Sehfähigkeit* entspricht. Diese Zusammenhänge werden durch einen Vergleich der durchschnittlichen *Differenzen von Verbal-IQ minus Handlungs-IQ* in den Visusgruppen I bis III nach Klauer (a.a.O., S. 576) noch unterstrichen: I = + 2,1 IQ; II = +15,0 IQ; III = +34,4 IQ; Gesamt = +13,3 IQ. Dagegen bleibt über die einzelnen Visusgruppen hinweg der *Gesamt-IQ relativ konstant*. Aus diesen Befunden leitete Klauer nun folgende Interpretationshypothesen ab:

1. Sehschwäche bewirkt im Regelfalle eine Herabsetzung des Handlungs-IQ bei gleichzeitiger Erhöhung des Verbal-IQ. Die Differenz zwischen Verbal- und Handlungs-IQ wird mit ausgeprägterer Sehschwäche positiv und größer (S. 576 ff.).

<sup>59</sup> Verbal-IQ (M = 100,8; s = 11,6), Handlungs-IQ (M = 101,2; s = 13,5), Gesamt-IQ (M = 100,7; s = 12,5). – Vgl. Priester 1958, S. 105.

2. "Bezüglich des Gesamt-IQ kompensiert der höhere Verbal-IQ die Beeinträchtigung beim Handlungs-IQ." (S. 575)
3. Die intellektuelle Streuung, besonders im verbalen Bereich, ist bei gegebener Sehbehinderung erheblich größer als im Regelfalle (Schultypenvergleich).
4. "Die ganzheitliche Auswirkung der Sehschwäche auf die intellektuelle Entwicklung (darf) als erwiesen gelten. Mit der *Beeinträchtigung der anschaulich-praktischen Intelligenz* geht eine *Akzentuierung der sprachlich-begrifflichen Intelligenz* einher . . . Es ist daher anzunehmen, daß sehbehinderte Kinder mit dem Grade ihrer Sehschwäche zunehmend *Techniken der verbal-intellektuellen Durchdringung der Wirklichkeit* und der *sprachlichen Orientierung entwickeln*." (S. 583)

Diese Thesen sind nicht unwidersprochen geblieben. Vor allem wandten sich Hudelmayer (1970, S. 58 ff.) sowie Baitinger & Bernd (1970) gegen zu weitreichende Schlußfolgerungen. Zwar wird die *Möglichkeit* "qualitativer Verlagerungen in den intellektuellen Fähigkeiten bei Sehschädigungen" eingeräumt, diese jedoch noch nicht als erwiesen angesehen. Vor allem wird die Klauersche These von der Steigerung der verbalen Intelligenz mit zunehmender Sehschädigung in Zweifel gezogen. Ferner werden Bedenken gegen die *Validität des HAWIK-Handlungsteils* in der Anwendung bei Sehbehinderten und daraus resultierende Generalisationen angemeldet. So schreibt Hudelmayer (S. 61 f.):

"Zu bezweifeln ist die Validität des HAWIK-Handlungsteils für die Entscheidung der Frage, ob bei Sehbehinderung (wenigstens bei einem Fernvisus von 1/4 und weniger) von einem 'Ausfall im Bereich der anschaulich-praktischen Intelligenz' gesprochen werden kann. Ein Festhalten an der operationalen Definition der Handlungs-Intelligenz in Form . . . des HAWIK-Handlungsteils halten wir für unangemessen, da dieser neben der Intelligenz die organische Sehfähigkeit – und zwar bei zunehmender Sehbehinderung in steigendem Maße – mitmißt und so das Ergebnis der intendierten Messung verfälscht; Blinde müßten dann konsequenterweise als hochgradig schwachsinnig hinsichtlich ihrer lebenspraktischen Intelligenz bezeichnet werden . . .

Eine zunehmende Verlagerung der Intelligenz bei zunehmender Sehschädigung zum Begrifflich-Sprachlichen hin halten wir nach dem von Baitinger und Bernd an 73 sehbehinderten und von Hudelmayer an 62 blinden und hochgradig sehbehinderten Kindern gewonnenen Material noch nicht für nachgewiesen. Die sehbehinderten und blinden Kinder unterscheiden sich in ihrem mittleren Verbal-IQ praktisch nicht voneinander. Viel deutlicher wirkte sich z.B. ihre Schulzugehörigkeit aus. Eine negative Korrelation zwischen Verbal-IQ und Sehfähigkeit, d.h. ein Ansteigen des Verbal-IQ bei Absinken der Sehfähigkeit, war nicht festzustellen. Demnach können wir auch eine 'große Diskrepanz zwischen den verbalen und den nicht-verbalen Intelligenzleistungen' nicht bestätigen."

Die Meinungen über mehr oder weniger identische operationale Untersuchungsbefunde sind also kontrovers. Für die Hudelmayersche Argumentation sprechen die niedrigen Korrelationen zwischen Sehfähigkeit und Intelligenz, besonders im Hinblick auf die Beziehung des *Fernvisus* (dem von Klauer ausschließlich gewählten Visuskriterium) zum HAWIK-IQ<sup>60</sup>, wohingegen die Korrelationen zwischen der Schulleistung in Rechnen, Lesen, Rechtschreiben und der HAWIK-Intelligenz durchweg Koeffizienten um 0.6 und 0.7 erbrachten – bei einer einzigen Ausnahme in Zeichnen/Formen, das nur mit  $r = 0.21$  (Blinde) resp.  $r = 0.51$  (Sehbehinderte) mit dem Verbal-IQ resp. Handlungs-IQ korrelierte. Ferner scheinen die HAWIK-Resultate der Stichproben von Baitinger & Bernd die "Scheren"-Hypothese Klauers, soweit sie den Verbal-IQ betrifft (vgl. Tab. 23),

<sup>60</sup> Baitinger u. Bernd (1970, S. 141) geben folgende Produkt-Moment-Korrelationen an für *Blinde*: "Visus" – Verbal-IQ  $r = -0.017$  bzw. für  
*Sehbehinderte*: Nahvisus – Verbal-IQ  $r = 0.33$   
Nahvisus – Handl.-IQ  $r = 0.44$   
Fernvisus – Verbal-IQ  $r = 0.17$   
Fernvisus – Handl.-IQ  $r = 0.17$

nicht zu bestätigen. Entsprechende Befunde Klauers müssen deshalb möglicherweise als Stichprobeneffekte interpretiert werden. Diese Deutung gewinnt umso größere Wahrscheinlichkeit, als Klauer lediglich den Fernvisus und nicht den für die HAWIK-Leistung wesentlich relevanteren Nahvisus<sup>60</sup> kontrollierte.

Trotz der z.T. berechtigten Kritik Hudelmayers an der Validität vorliegender Aussagen in bezug auf den HAWIK-Handlungsteil kann ohne größeres Risiko im Sinne Klauers festgehalten werden, daß die "Handlungsintelligenz" mit zunehmender Sehschädigung absinkt. Die operationalen Befunde sind hier u.E. zwingend und zeigen unmißverständlich bestimmte Zusammenhänge, die freilich nicht überinterpretiert werden dürfen, an. Dagegen ließ sich die Kompensationshypothese Klauers in den nachfolgenden Untersuchungen nicht eindeutig verifizieren. Immerhin stimmen aber die HAWIK-Profile Sehgeschädigter in den charakteristischen Merkmalsdimensionen relativ gut überein (vgl. Grafiken 12-14).

Demnach weisen *Sehbehinderte* intra- und teilweise auch interindividuell überdurchschnittlich hohe Leistungen im Allgemeinen Verständnis, Gemeinsamkeitenfinden und Wortschatztest (Verbalteil) sowie im Mosaiktest (Handlungsteil) auf, während *Blinde* ihre eindeutigen Schwerpunkte im Gemeinsamkeitenfinden, Wortschatztest und Zahlennachsprechen haben. Dagegen sind die Leistungen Sehgeschädigter im Allgemeinen Wissen und Rechnerischen Denken (Verbalteil) vergleichsweise niedrig, während sämtliche Subtestleistungen im Handlungsteil des HAWIK – mit der schon genannten Ausnahme des Mosaiktests – in ihrem Niveau erheblich unter der Norm Sehender liegen. Diese Aussagen gelten vorwiegend für die blinde resp. sehbehinderte *Volksschulpopulation*. Für die Gruppe der *lernbehinderten* Blinden versus der blinden *Mittel- und Oberschüler* ergeben sich demgegenüber gewisse Akzentverschiebungen im HAWIK-Profil, wie eine Hamburger Untersuchung an insgesamt 46 blinden und hochgradig sehbehinderten Schülern zeigt. So ermittelte Rath in den Jahren 1964-66 folgende Mittelwerte der verbalen Untertests im HAWIK (s. Tab. 24).

Nach unserer Tabellenübersicht liegen die durchschnittlichen Intelligenzleistungen *lernbehinderter* Blinder in der Regel unter denen der sehtüchtigen Lernbehinderten, besonders in den für die Lernbehindertendiagnostik (Auslese) diakritischen Funktionen von RD, AW und ZN (Schulleistungsfaktor). Ausnahmen bilden die Leistungen im Wortschatztest und Gemeinsamkeitenfinden, wo die Blindengruppe etwas besser ist. Sämtliche M-Differenzen sind jedoch nicht signifikant und dürfen somit nur als Tendenzen in der jeweils aufgezeigten Richtung interpretiert werden. Umgekehrt erweisen sich die HAWIK-Leistungen der sehgeschädigten *Real- und Oberschüler* im Allgemeinen Wissen, Rechnerischen Denken, Wortschatztest und Zahlennachsprechen (hier signifikant) denen der sehenden Vergleichsgruppe überlegen, während die Dimensionen Allgemeines Verständnis und Gemeinsamkeitenfinden zugunsten der Sehenden diesmal ausfallen – bei insgesamt etwas erhöhtem Level der Blinden. Damit wäre indirekt ein Zusammenhang zu den von Hayes gefundenen Verteilungsproportionen (s.S. 207; vgl. auch Tab. 25) hergestellt. Ferner korrelierte Rath die Ergebnisse im HAWIK-Verbalteil mit denen im IPT und fand Koeffizienten von 0.57 (Mittel- und Oberschüler) bzw. 0.69 (Hilfsschüler); die abnehmenden Zusammenhangswerte bei erhöhtem Schul- bzw. IQ-Niveau interpretierte die Autorin im Sinne Wewetzers Divergenzhypothese. Ähnliche Resultate fand übrigens Priester in der Beziehung HAWIK-Verbalteil und SIT (Stanford-Intelligenztest), nämlich bei *sehenden* Oberschülern  $r = 0.60$  sowie Lernbehinderten  $r = 0.84$ .

In den aufgewiesenen Rahmen intelligenzdiagnostischer Merkmalsdimensionen Sehgeschädigter paßt eine Reihe weiterer bekannt gewordener Untersuchungsbefunde mehr oder weniger stimmig. So faktorenanalytierte Tillman 1967 die verbalen Intelligenzleistungen nach WISC von jeweils 110 blinden und sehenden Schülern im Alter von 7 bis 12 Jahren und fand:

"Across groups, the obtained factor structures were similar on Arithmetic, Information,



Tabelle 24

Mittelwerte blinder und hochgradig sehbehinderter Hilfs-, Mittel- und Oberschüler in den Subtests des HAWIK-Verbalteils im Vergleich zu den entsprechenden Kennwerten der HAWIK-Standardisierungstichproben Sehender (vgl. Rath 1967, S. 126 f.)

Subtest	Blinde Lbh. (N = 23)	Sehende Lbh. (N = 150) <sup>61</sup>	Blinde Real/ Oberschüler (N = 23)	Sehende Real/ Oberschüler (N = 150) <sup>61</sup>
AW	4,55 <sup>62</sup>	5,42	13,00	12,53
AV	4,55	5,28	11,52	12,40
RD	3,55	5,29	12,57	11,94
GF	6,21	5,22	11,17	11,72
WT	5,87	5,20	12,87	12,63
ZN	5,42	6,40	15,00	11,66
Profil- durchschn.	5,03	5,47	12,69	12,15

and Comprehension but less so on Similarities and Vocabulary. Fewer factor loadings and weaker communalities for the blind on all subtests except Arithmetic suggest a greater specificity in the organization of abilities by these tests." (S. 112)

Der Autor folgert hieraus, daß blinde Kinder ihre unterrichtlichen Erfahrungen weniger gut zu integrieren vermögen, was zu isolierten Wissensdaten führe, deren Einordnung in ein Bezugssystem erschwert sei. Ihre verbalen Fähigkeiten basierten auf einem wenig differenzierten (elaborierten) Grundwortschatz. Auch neigten blinde Kinder dazu, Abstraktionsprobleme begrifflicher Art (Konzeptbildungen) auf niedrigerer, d.h. konkreter bzw. funktionaler, Stufe zu lösen, was zu Retardierungen gegenüber der sehenden Vergleichsgruppe führe. Dagegen seien ihre rechnerischen Leistungen denen der Sehtüchtigen äquivalent. Weiterhin fand Juurmaa (1967), der 125 blinde Kinder und Jugendliche hinsichtlich der Intelligenzleistungsdimensionen Verbal Comprehension, Numerical Ability und Arithmetic Reasoning mit einer sehenden Gruppe verglich, *keine signifikanten* Differenzen. Somit sah er sich in der Hypothese bestärkt, daß intellektuelle Funktionen, soweit sie keine visuellen Vorstellungen voraussetzen, durch Ausfallserscheinungen des Gesichtssinnes nicht beeinträchtigt werden würden.

Eine soeben von Hengstler (1972) vorgelegte Arbeit, die es verdiente fortgesetzt zu werden kommt zu folgendem interessanten Resümee.

"Bei der Untersuchung der Intelligenzfaktoren der HAWIK-Leistung Sehgeschädigter schien sich zu zeigen, daß die Intelligenzstruktur *Sehbehinderter* Faktoren offenbart, die mit den bei Schwachbegabten ermittelten identisch sein könnten. Außerdem ergab sich, daß die Anwesenheit eines Gruppenfaktors mit dem Prädikat "Handlung" im Sinne Wechslers, welcher bei Sehgesunden nachgewiesen ist, zumindest bei einigen Sehbehindertengruppen in Zweifel gezogen werden muß. Stattdessen scheinen unter anderem Gruppenfaktoren wirksam zu sein, die die Identifikation und Wiedergabe von Strichsymbolen bzw. Linienzeichnungen einerseits und die Organisation von Flächenteilen im Raum andererseits ermöglichen. Ebenfalls scheint die Gruppe der Aufmerksamkeits- und Lernfaktoren ("Trace" u.a.) in einem größeren Zusammenhang mit einem sekundären Faktor ("Umwelt" zu stehen, der sich in eine gezielte Lern-Umwelt-Komponente und in eine mehr locker-"zufällige" Situationskomponente polarisiert.

Bei *Blinden mit Sehrest* zeigt sich – da der Handlungsteil und somit die Handlungsintelligenz hier nicht getestet werden kann – ebenfalls diese "Umweltbezogenheit", die sich

<sup>61</sup> Nach Priester, a.a.O., S. 119.

<sup>62</sup> Wertpunkte (WP)

wiederum in "gezielt" und "nicht gezielt" polarisiert. *Blinde ohne Sehrest* scheinen ihre Verbalintelligenz in den beiden Komponenten "Gedächtnis" und "Einordnung" der Eindrücke von außen zu manifestieren.

Im ganzen dürfte die Intelligenzstruktur Sehgeschädigter hauptsächlich hinsichtlich der Handlungsintelligenz anders gelagert sein als die Sehgesunder, aber auch die Verbalintelligenz scheint eine Polarisierung zu bevorzugen" (S. 41).

Nach Überprüfung – und Ablehnung – der Nullhypothese kann Hengstler (loc. cit.) schließlich die *Abhängigkeit der Intelligenzstruktur* Sehgeschädigter (Sehbehinderter und Blinder) von folgenden *Variablen* belegen: Fernvisus bzw. Nahvisus versus Blindheitsgrad, Schuljahr und Schulart (Schulbesuchsform). Je nach Variation der genannten Bedingungen ist demnach mit *verschiedenen* Faktorenstrukturen bei sehgeschädigten Kindern und Jugendlichen in der HAWIK-Intelligenz(leistung) zu rechnen. Auf die psychodiagnostischen Implikationen solcher Befunde kann hier nur verweisend aufmerksam gemacht werden (siehe besonders S. 120ff. u. 185ff.).

Was endlich den Leistungsvergleich zwischen Blinden und Sehbehinderten betrifft, so erbrachten einschlägige Untersuchungen teils bessere (Wappmann), teils schlechtere (Strehle, Rexhausen, Schweingruber, Köhn, Schirmer) *verbale* Intelligenzleistungen Sehbehinderter. Dalferth (1964) stellte vor allem die schlechteren *Gedächtnisleistungen* sehschwacher Schüler heraus. Die genannten Differenzen waren allerdings selten signifikant zu machen. Hingegen fanden Hudelmayer und Baitinger u. Bernd, abgesehen von Leistungsvorteilen der Sehbehinderten in AV und der Blinden in ZN, keine auffälligen Differenzen der beiden Vergleichsgruppen. Freilich muß bei allen bisher vorliegenden profilanalytischen Interpretationsversuchen die bereits mehrfach erörterte Testkritik am WISC bzw. HAWIK (vgl. besonders S. 120ff. u. S. 185ff. in diesem Buch) als Kautel hier noch angefügt werden.

Bevor wir abschließend einige Fragen des sozio-kulturellen Kontextes erörtern, seien in der folgenden Tabelle 25 die Verteilungsparameter der baden-württembergischen Blinden- und Sehbehinderten-Populationen denjenigen der HAWIK-Standardisierungsstichprobe (Sehende) gegenübergestellt. Die diesbezüglichen Ergebnisse Klauers und Hayes' (34% bei iHB bzw. 43,5% bei WAIS-Blinde gegenüber den mittleren 50% bei Terman bzw. Wechsler) können somit voll bestätigt werden.

Tabelle 25  
IQ-Verteilung Sehgeschädigter im Vergleich zur Normalklassifizierung im HAWIK  
(nach Baitinger & Bernd 1970, S. 139)

HAWIK-IQ	Sehende HAWIK	Blinde Verbal	Verbal	Sehbehinderte Handl.	Gesamt
	%	%	%	%	%
130 u. mehr	2,2	6,5	9,6	1,4	2,7
120-129	6,7	20,9	16,4	2,7	6,8
110-119	16,1	17,7	9,6	8,2	12,3
90-109	50,0	30,6	49,3	36,9	43,8
80- 89	16,1	12,9	8,2	17,8	19,2
70- 79	6,7	6,5	6,8	19,2	8,2
69 u. weniger	2,2	4,8	–	13,7	6,8

Die vorstehenden Tabelleninformationen möchte ich vor allem unter zwei Gesichtspunkten interpretieren. Einmal fallen die hohen Anteile Blinder und Sehbehinderter in den oberen IQ-Bereichen des HAWIK-Verbalteils auf, die sich u.E. nur plausibel mit Hilfe der Klauer-schen *Kompensationshypothese* (s.S. 212f.) interpretieren lassen. Da Hayes u.a. zu ähnlichen operationalen Befunden gelangten, kommen Stichprobenfehler oder Methodenartefakte hierfür kaum in Betracht, d.h. die beobachteten Verteilungsdifferenzen müssen als charakteristisch für Sehgeschädigte angesehen werden. Zum andern stehen die gesteigerten Frequenzanteile Blinder (im HAWIK-Verbalteil) und Sehbehinderter (besonders im HAWIK-Handlungsteil) in den unteren IQ-Bereichen nach meinem Dafürhalten zuallererst direkt oder indirekt – in Zusammenhang mit einer Mehrfachschädigung. So fand etwa Schirmer (1971, S. 78 ff.) im Rahmen ihrer Wortschatztests bei *Sehbehinderten*, daß 15% (N = 83) ihrer Untersuchungspopulation (N = 551) *Zusatzbehinderungen* aufwiesen, 1,5% (N = 8) *sogar zwei zusätzliche Behinderungen* (neben der Sehschädigung). Hiervon entfielen – die Mehrfachbehinderungen jeweils mitgerechnet – 5% auf Zerebralschäden, 4% auf Sprachstörungen, 2,2% auf Schwerhörigkeit, 1,6% auf spastische Störungen und knapp 3% auf sonstige Behinderungsarten, z.B. Gehbehinderung, Hüftluxation, psychische Störungen, Wachstumsstörungen, Herzkrankheiten etc. Die Vermutung, daß *Mehrfachbehinderungen* bei sehschwachen Schülern in der Regel, d.h. überzufällig häufig, Intelligenzminderungen (zumindest in den Dimensionen Verbal Comprehension und Verbal Reasoning) zur Folge haben, wird nun entscheidend durch zwei Tatsachen gestützt: 1 WST-Leistungsvergleiche zwischen Kindern mit (N = 83) und ohne (N = 468) zusätzliche Behinderungen neben der Sehschädigung erbrachten *sehr signifikante Unterschiede zuungunsten der Mehrfachgeschädigten*; 2 Schirmer ermittelte für ihre Mehrfachgeschädigtengruppe folgende Verteilungsparameter (in Klammern die Frequenzanteile der "Nur"-Sehbehinderten): 65% (47,5%) unter 100 IQ versus 35% (52,5%) über 100 IQ oder 36% (12%) unter 85 IQ versus 15,5% (11%) über 115 IQ. Die sehr signifikanten Unterschiede gehen eindeutig zu Lasten der WST-Diskrepanzen beider Gruppen in den *unteren* IQ-Bereichen, womit die *intelligenzmindernde Funktion der Mehrfachbehinderung* als Interpretationshypothese hinreichend empirisch verifiziert wäre. Dagegen konnten Jordan & Felty (1968) bei amerikanischen Untersuchungen an 253 sehschwachen Schülern keine signifikanten Korrelationen zwischen der Sehschädigung allein (Grad der Sehschwäche) und dem mittleren Intelligenzniveau feststellen, was sich gut mit den Ergebnissen von Hudelmayer et al. deckt. Ebenso wenig bestätigten sich korrelative Beziehungen zwischen dem Zeitpunkt der Erblindung (prä- versus postnatal) und allgemeinem Intelligenzniveau (Hayes 1941). Wohl aber fand Schirmer (a.a.O., S. 71 ff.) signifikante, wenngleich niedrige Zusammenhänge zwischen der Sehschärfe und der WST-Bearbeitungszeit versus (nur) beim WST 7-8 sehr signifikante, mittlere Korrelationen zwischen der Länge der Testbearbeitungszeit und der Höhe der Testleistungen.

Am Ende unseres kurzen Übersichtsreferates soll – wie oben schon angekündigt – noch kurz auf mögliche Auswirkungen des sozio-kulturellen Umfeldes im Hinblick auf die intellektuelle Entwicklung Sehgeschädigter eingegangen werden. Sowohl Jordan & Felty als auch Schirmer konnten bei ihren Untersuchungsgruppen Sehbehinderter *keine* signifikanten Zusammenhänge zwischen sozioökonomischem Status des Elternhauses (Vaterberuf und Intelligenztestleistung nachweisen. Offensichtlich – so meinen die Autoren – wirken sich Sehschädigung so gravierend auf das kognitive Verhalten aus, daß demgegenüber Einflüsse der sozialen Schichtzugehörigkeit, des sozialkulturellen Familienmilieus usw. relativ bedeutungslos werden. Weiterhin schwäche das institutionelle Klima der Sonderschule, besonders im Internatsverbund, etwaige (frühere) familiäre Einflüsse erheblich ab und wirke nivellierend auf außerschulische Entwicklungs- und Bildungsanreize. Übereinstimmend dazu fand Schirmer beim Vergleich sehbehinderter Schüler des Internats versus Externats in bezug auf die angesprochenen Milieuvariablen keine signifikanten Beziehungen zur (verbalen) Intelligenztestleistung.

Während Klauer sowie Baitinger & Bernd bei ihren Untersuchungen Sehbehinderter keine Unterschiede in der Intelligenzleistung jüngerer und älterer Schüler feststellen

konnten, variiert nach den Befunden von Jordan & Felty die Schulbesuchszeit (Dauer der Sonderbeschulung?) mit der Höhe des WISC-IQ. Nach Ansicht der Autoren kann das Anwachsen des IQ im Laufe der Schuljahre mit formaler Erziehung, informellen sozialen Erfahrungen oder/und verzögerter Reifung des Intellekts zu Beginn der Schulzeit zusammenhängen. Unabhängig von der Organisationsform der Bildungseinrichtungen für Blinde und Sehbehinderte (Sonderschule oder Gesamtschule) muß zum Schluß festgehalten werden, daß hier der *Schule* die Hauptlast der intellektuellen Förderung zufällt. Möglicherweise gilt diese Forderung generell im Bereich der Sondererziehung, sicher in verstärktem Maße für die Gruppe der Mehrfachbehinderten.

## IV. Desiderata

Im empirischen Teil des Buches konnten zahlreiche – überwiegend neuere – Untersuchungsergebnisse zur Begabungsdiagnostik in der Schule und Sonderpädagogik erörtert werden, ohne daß Vollständigkeit bezüglich der Beiträge pädagogisch-psychologischer Provenienz angestrebt oder gar erzielt worden wäre. Ich meine aber, daß die angesprochenen Probleme und Forschungsinhalte die zentralen Themen intelligenzdiagnostischer Anwendung in Bildung und Erziehung hinreichend repräsentieren. Im Hinblick auf die pädagogische Praxis, in deren Dienst letzten Endes alle diagnostischen Bemühungen stehen (siehe unsere früheren Darlegungen auf S. 81ff.), zeichnen sich nunmehr – neben dem Auftrag der praktischen Verwertung gesicherter Untersuchungsergebnisse – folgende Punkte als vordringliche Aufgabenziele ab: 1) eine allgemeine Verbesserung der Meßmethoden (Intelligenztests) für den Bildungsbereich, 2) eine adäquatere Berücksichtigung sonderpädagogischer Bedürfnisse und Fragestellungen, 3) die Objektivierung der Lehrerurteile über Schulleistung(en) und Lernfähigkeit(en) qua Bezugskriterien intelligenzdiagnostischer versus schuleignungsprognostischer (begabungsdifferenzierender) Aussagemöglichkeiten.

Ad 1): Die Forderung nach besseren, d.h. für die schulischen und pädagogischen Belange *brauchbareren Intelligenztests* wird in letzter Zeit häufig erhoben und impliziert verschiedene Problemaspekte. So wird etwa das sprachliche Übergewicht vieler Intelligenztests beklagt, bei deren Verwendung man eine Benachteiligung zumindest für die Kinder und Jugendlichen befürchtet, die den "restringierten" Sprachkode verwenden. Oder es werden mangelnde Testgütekriterien einzelner Verfahren – zu Recht – moniert. Vielfach werden dann überhaupt Intelligenztests als Meßinstrumente kognitiver Fähigkeiten abgelehnt, wobei die betr. Kritiker nicht selten einseitige und somit unhaltbare theoretische Positionen bezüglich der Phänomene Begabung und Begaben einnehmen. Wie immer man zur Leistungsdiagnostik allgemein oder speziell zur Intelligenzdiagnostik im Bildungsbereich stehen mag, die Notwendigkeit *methodischer* Verbesserungen dürfte dem aufmerksamen Leser dieses Buches kaum entgangen sein. So fehlt es vielfach an der Repräsentanz testdiagnostischer Aussagen, d.h. die verwendeten Testitems oder erfaßten Variablen repräsentieren in solchen Fällen nicht die erforderliche Aspektfülle im Hinblick auf ein bestimmtes Untersuchungsziel, z.B. die Funktion der Intelligenz beim Zustandekommen versus unerwarteten Defizit bestimmter Schulleistungen, die Variablenstruktur kognitiver Fähigkeiten in bezug auf diese oder jene Schuleignung, Einflüsse von Sinnesschäden (Hörschädigung, Sehschädigung u.ä.) auf die intellektuelle Lernleistungsfähigkeit u.dgl.m. Oder es werden bestimmte Eigenschaften, z.B. Kreativität, weder im Testdesign noch in Schul- und Bildungskonzepten angemessen berücksichtigt (vgl. hierzu Ulmann 1968, Massialas & Zevin 1969, Mühle & Schell 1970 u.a.).

Im Bildungsbereich eingesetzte bzw. pädagogisch brauchbare Intelligenzmeßverfahren (Tests) müssen somit grundsätzlich folgenden Anforderungskriterien genügen: a) Die Tests müssen *objektiv* durchführbar und *eindeutig* auswertbar sowie die Zuverlässigkeit (*Reliabilität*) und Gültigkeit (*Validität*) ihrer Aussagen (Testergebnisse) im jeweiligen schulischen und pädagogischen Anwendungsfeld kontrolliert und für tauglich befunden worden sein. Insbesondere ist der Frage der Validität von Intelligenztests im Hinblick auf den jeweiligen Untersuchungszweck (z.B. Schuleignungsprognose) größte Aufmerksamkeit zu widmen.

b) In engem Zusammenhang mit dem Validitätsanspruch steht die Forderung nach der *Itemrepräsentanz*, d.h. die im Test verwendeten Aufgaben(reihen) sollen in bestmöglichen Umfange die – idealiter alle – in bezug auf den Untersuchungsgegenstand relevanten Faktoren abdecken. Für Zwecke der Schuleignungsermittlung beispielsweise impliziert diese Forderung die Berücksichtigung sowohl der sog. Kernintelligenz (Denken, verbale und nonverbale Intelligenzfaktoren, mathematische und technisch-konstruktive Fähigkeiten) als auch für den Schul- und Bildungserfolg einflußreicher Hilfs- und Stützfunk-

tionen (Gedächtnis, Bildungsinteressen, Lern- und Leistungsmotivation u.ä.). Am ehesten werden diesen Zielen erfahrungsgemäß heterogene Testbatterien, also faktorisierte Tests in gemischter Zusammenstellung, gerecht.

c) In der Schule benutzte Tests sollten mit besonderer Sorgfalt *altersgemäß*, d.h. dem jeweiligen Entwicklungsstand der Kinder und Jugendlichen angepaßt, sowie in Form und Inhalt der Aufgaben für Schüler (Probanden) und Lehrer (Testleiter) gleichermaßen ansprechend sein. Wenn darüberhinaus Gesichtspunkte der *Ökonomie* (niedrige Kosten, geringer Zeitaufwand, automatische Auswertung) verwirklicht werden konnten, handelt es sich fast schon um ideale Voraussetzungen (intelligenz)testdiagnostischer Untersuchungen vorhaben. Gleichwohl bleibt festzuhalten, daß viele der hier angesprochenen Punkte vorläufig allenfalls programmatischer Natur sind und der Initiative künftiger Testautoren überantwortet werden müssen.

Ad 2): Im Kontext *sonderpädagogischer* Fragestellung gelten im Prinzip dieselben Forderungen, hinzu kommen jedoch hier einige nicht unbedeutende Akzentverschiebungen. Bislang fehlt es noch weithin an den notwendigsten Testverfahren für die verschiedensten sonderpädagogischen Zielgruppen (Lern- und Geistigbehinderten, Körperbehinderten, Sprachbehinderten, Hör- und Sehgeschädigten usw.). Häufig werden deshalb mehr oder weniger unkontrolliert gängige Tests, die im Hinblick auf den Regelfall konstruiert wurden, ohne weiteres für sonderpädagogisch-diagnostische Zwecke (z.B. Klassifikationsaufgaben, unterrichtliche Differenzierungsprobleme u.ä.) übernommen, wobei sich das Fehlen populationspezifischer Testnormen vielleicht am unangenehmsten bemerkbar macht. Die Adaptation solcher Verfahren sowie ihre teststatistische Absicherung unter sonderpädagogischen Anforderungskriterien bereitet freilich größere Schwierigkeiten, als es Außenstehenden erscheinen mag. Die Gründe dafür liegen einmal in der ausgeprägteren Gruppenheterogenität bzw. Komplexität der Problemfälle und dem häufig damit korrespondierenden Informationsdefizit bezüglich Erscheinungsform(en), Deskriptionskriterien, Ätiologie usw., zum andern aber sind die Ursachen für entsprechende Informationsmängel mehr äußerer Natur, etwa durch geringe Gruppengrößen bzw. disparate Auftretenshäufigkeiten bei extensiver regionaler Streuung bedingt, wobei der Informationszugang von vornherein erschwert wird. Schließlich müssen auch man-power-bedingte Informationsrückstände im Bereich sonderpädagogischer Forschung in diesem Zusammenhang veranschlagt werden, insofern sich nämlich nur eine begrenzte Anzahl engagierter Forscher den *sonderpädagogischen* Disziplinen und ihren Spezialproblemen zuwendet. Das Heer der Forscher auf den von allgemeinem Interesse begleiteten Untersuchungsfeldern ist nach wie vor – absolut und relativ betrachtet – ungleich größer. In gewisser Weise kehren sich hier die Proportionen um: Je schwieriger und komplexer und damit arbeitsaufwendiger die Forschungsaufgaben werden, desto weniger Forscher (und häufig auch weniger Mittel) stehen zur Verfügung. Dabei spielen unterschiedliche Größenverhältnisse in den Zielpopulationen (hier Anzahl der Kinder und Jugendlichen in der Regelschule versus in der Sonderschule) im Hinblick auf den tatsächlichen Forschungsaufwand – entgegen weitverbreiteten Ansichten – nur eine untergeordnete Rolle: Eine intelligenzdiagnostische Untersuchung von 100 über das ganze Land verstreut wohnenden blinden Schülern der 4. Grundschulklasse (Population) erfordert beispielsweise keinen geringeren, eher größeren arbeitszeitlichen Aufwand als eine entsprechende Testerhebung bei 1000 normalsinnigen Viertklässkindern (Stichprobe) einer bestimmten Wohnregion. Dieses Beispiel ließe sich leicht fortsetzen. Die Konstruktion sonderpädagogisch geeigneter Untersuchungsverfahren oder auch die Adaptation einzelner für die Normalpopulation(en) konzipierter Tests für spezifische Belange Behinderter wird – wie ähnliche Aufgaben – nur dann befriedigend zu leisten sein, wenn man an verantwortlicher Stelle (Ministerien, Forschungsinstitutionen, Fördergesellschaften u.ä.) zu *sachlich* angemessenen Lösungen, z.B. problemorientierten Verteilungsschlüsseln bei der Mittelvergabe, bereit ist. Ebenso müßten die personellen Voraussetzungen, vorab im Hinblick auf eine effizientere Arbeitsteilung in Forschungsteams, neu durchdacht werden.

Neben diesen generellen Verbesserungsvorschlägen erhebt sich eine Reihe konkreter Methodenwünsche, auf die bereits in früheren Erörterungen hingewiesen wurde. Zweifellos führen hier Themen wie Neukonstruktion brauchbarer Meßverfahren oder Bearbeitung bereits vorhandener, für die Anwendungspraxis im sonderpädagogischen Feld jedoch unzulänglicher, Tests eine solche Wunschliste an. Für die Gruppen der Hör- und Sprachbehinderten sowie die der Blinden und Sehbehinderten wären demnach weitere *nonverbal* Intelligenztests zu erproben, z.B.: die "Leiter International Performance Scale" von Leiter (1969), der "Figure Reasoning Test" von Daniels (1962) oder die "Bildertest"-Serien BT 1-2 (vgl. Horn & Schwarz 1967) und BT 2-3 (vgl. Ingenkamp 1966). Ferner wären Tests zur Messung *verbaler* Fähigkeiten zu erstellen resp. zu adaptieren: Wortschatztests, Fremdspracheneignungstests, Meßskalen zur Analyse grammatischer und syntaktischer Sprachstrukturen u.ä. (siehe oben S. 52 ff.). Für andere Behindertengruppen sind wiederum teilweise dieselben, teilweise andere Intelligenzmeßverfahren von Interesse (vgl. z.B. Zimmermann 1969). Die von uns im Anhang dieses Buches beigelegten *Testnormen* für Hörgeschädigte wären in bezug auf die zugrundeliegenden Stichprobengrößen (teilweise) zu vervollständigen und durch weitere Verfahren zu ergänzen. Analog müßten *Testnormen* für die übrigen Behindertengruppen erstellt oder/und Testbearbeitungen versus Neukonstruktionen in Angriff genommen werden. Bereits vorliegende Befunde sollten gesammelt und der Öffentlichkeit bzw. Praxis zugänglich gemacht werden, selbst wenn den Ergebnissen nur vorläufiger Charakter eignet. Jedenfalls sehe ich momentan keine andere Möglichkeit, wissenschaftliche Ergebnisse (die mehr oder minder immer begrenzte bzw. bedingte Gültigkeit haben werden – freilich auch jederzeit überprüfbar sein müssen) einerseits und praktische Notwendigkeiten andererseits fruchtbar miteinander zu verbinden.

Ad 3): Die Frage einer *objektivierten Leistungsbeurteilung* in der Schule steht mit Problemen der Intelligenzbestimmung nur in mittelbarem, nichtsdestoweniger bedeutsamem Zusammenhang. Einerseits müssen nämlich – darauf wurde wiederholt hingewiesen – intellektuelle und Persönlichkeitsfaktoren sowie effektive versus defiziente (mangelnde) Schulleistungen in einem gewissen Abhängigkeitsverhältnis zueinander gesehen werden; nur unter dieser Voraussetzung bekommen Intelligenzuntersuchungen in der Schule vielfach erst ihren Sinn, wenigstens sofern man die Zielthematik (Tests als Hilfen im Erziehungs- und Bildungsprozeß) nicht außer acht läßt. Zum andern aber interessieren hier Schulleistungsvariablen qua Kriterien des Schul- und Bildungserfolges und damit als unverzichtbare "Außenkriterien" für die Gültigkeitsbestimmung diagnostischer versus prognostischer Urteile jedweder Provenienz. Ohne diesen Bezug zur Schulleistung und zum Bildungserfolg wären Intelligenztestaussagen im Bereich schulischer Anwendung zumeist wenig brauchbar, weil entsprechende Kontrollmöglichkeiten (Veri- oder Falsifikation testdiagnostischer Aussagen) fehlten, die die Verbindlichkeit des Testurteils erst garantieren.

Schließlich wird vom Schul- und Bildungssystem selbst gefordert werden müssen, daß es – neben der (optimalen) Ermöglichung von Schülerleistungen – unter objektiven Bedingungen gesicherte, d.h. *zuverlässige* und *gültige* Leistungsbewertungen zustandebringt. Nur objektive, reliable und valide Leistungsbeurteilungen innerhalb der Schule, beispielsweise meßtheoretisch einwandfreie Lehrerurteile (objektivierte Notengebung), garantieren "gerechte" Schülerzensuren. Unter gewissen Vorbehalten können diese als Indikator schulischen Bildungserfolgs überhaupt und somit auch als Kriterium intelligenzdiagnostischer Ziele angesprochen werden. Die Vorbehalte betreffen im besonderen die Tatsache, daß Schülerleistungen stets in einem hochkomplexen Interdependenzgeschehen zustande kommen; daraus abgeleitete "pädagogische" Folgerungen für die Familien- und Schulerziehung bzw. die intellektuelle Förderung des Kindes und Jugendlichen wurden andernorts ausführlich diskutiert (s.S. 47 ff.) und brauchen hier nicht mehr wiederholt zu werden.

Vieles von dem, was zuletzt angesprochen wurde, ist heute noch Desiderat. Es bedarf der Anstrengung aller Beteiligten, der Erzieher und Lehrer, der allgemeinen Schul- und Sonderpädagogen, der Testpsychologen und nicht zuletzt des Diagnostikers, die skizzierten Ziele intelligenzdiagnostischer Aufgabenfunktionen zu realisieren.

## **V. Tabellen-Anhang**



Tabelle I a Altersnormen zum Progressiven Matrizen-Test für hörgeschädigte Kinder und Jugendliche (Spaltenwerte = PMT-Rohpunkte)

IQ	5;0 bis 5;5	5;6 bis 5;11	6;0 bis 6;5	6;6 bis 6;11	7;0 bis 7;5	7;6 bis 7;11	8;0 bis 8;5	8;6 bis 8;11	9;0 bis 9;5	9;6 bis 9;11	PR
55											0,1
56,5											0,2
58											0,3
59,5				1	1	1	1	1	1	1	0,4
61	1	1	1	.	2	2	2	2	2	2	0,5
62,5				2	.	3	3	3	3	3	0,6
64	2	2	2	.	3	.	4	4	4	.	0,8
65,5	.	.	3	3	.	4	.	.	.	4	1
67	3	3	4	4	4	.	5	5	5	5	1,5
68,5	.	4	4	.	.	5	.	.	6	6	2
70	4	.	.	5	5	6	6	6	7	7	2,5
71,5		5	5	.	6	.	.	7	.	8	3
73	5	.	.	6	.	7	7	8	8	.	3,5
74,5	6	6	6	.	7	8	8	9	9	9	4
76	.	.	7	7	8	.	.	.	10	10	5
77,5	7	7	.	8	.	9	9	10	.	11	7
79	.	8	8	.	9	.	.	11	11	12	8
80,5	8	.	9	9	10	10	10	12	12	.	10
82	.	9	.	10	.	11	.	.	13	13	12
83,5	9	.	10	11	11	.	11	13	.	14	13
85	.	10	.	11	12	12	12	14	14	15	16
86,5	10	.	11	12	.	.	15	15	.	.	18
88	.	11	.	.	.	13	13	.	16	16	21
89,5	11	.	12	13	13	.	.	16	.	17	24
91	.	12	.	.	.	14	14	17	17	18	27
92,5	.	.	13	.	.	.	.	18	18	.	31
94	12	.	.	14	14	15	15	.	19	19	34
95,5	.	13	14	.	.	16	16	19	20	20	38
97	.	.	.	.	15	16	.	20	.	21	42
98,5	13	.	.	15	.	17	17	21	21	22	46
100	.	14	.	.	.	17	.	22	22	23	50
101,5	.	.	15	16	16	.	18	.	23	24	54
103	.	.	.	.	.	18	.	23	.	25	58
104,5	14	.	.	.	17	.	19	24	24	25	62
106	.	15	16	17	.	19	20	25	25	27	66
107,5	.	.	.	.	.	.	.	.	26	28	69
109	.	.	.	18	18	20	21	26	.	29	73
110,5	.	.	.	.	.	.	.	27	27	30	76
112	15	16	17	.	19	21	22	28	28	31	79
113,5	.	.	.	19	.	.	.	.	29	32	82
115	.	.	.	.	.	22	23	29	.	33	84
116,5	.	.	18	.	20	.	24	30	30	34	86
118	.	.	.	20	.	23	.	31	31	35	88
119,5	16	17	.	.	21	.	25	.	32	36	90
121	.	.	.	21	.	24	.	32	33	.	92
122,5	.	.	19	.	22	.	26	33	.	37	93
124	.	.	.	.	.	25	.	34	34	38	95
125,5	17	18	.	22	.	.	27	35	35	39	96
127	.	.	.	23	23	.	28	.	36	40	96,5
128,5	.	.	20	23	.	26	.	36	37	41	97
130	.	.	.	.	24	.	29	37	38	42	97,5
131,5		19		.	.	27	30	38	39	43	98
133	18	.	21	24	.	27	31	.	40	44	98,5
134,5	.	.	.	.	25	28	32	39	41	45	99
136	.	.	.	25	.	.	33	40	42	46	99,2
137,5	.	20	.	.	26	29	34	41	43	47	99,4
139	.	.	22	.	.	.	.	.	44	48	99,5
140,5	19	.	.	26	.	30	35	42	45	49	99,6
142	.	.	.	.	27	.	36	43	46	50	99,7
143,5	.	21	23	.	.	31	37	44	47	51	99,8
145	20	.	.	27	28	32	38	45	48	52	99,9

Fortsetzung Tabelle I a

IQ	10;0 bis 10;5	10;6 bis 10;11	11;0 bis 11;5	11;6 bis 11;11	12;0 bis 12;5	12;6 bis 12;11	13;0 bis 13;5	13;6 bis 13;11	14;0 bis 14;5	14;6 bis 14;11	PR
55					2	4	5	11	12	13	0,1
56,5					3	5	6	12	13	14	0,2
58				1	4	6	7	13	14	15	0,3
59,5	1	1	1	2	5	7	8	14	15	16	0,4
61	2	2	2	3	6	8	9	15	16	17	0,5
62,5	3	3	3	4	7	10	11	16	17	18	0,6
64	4	4	4	5	9	11	12	17	18	19	0,8
65,5	5	5	5	6	10	12	13	18	19	20	1
67	6	6	6	7	11	13	14	19	20	21	1,5
68,5	7	7	7	8	12	14	15	20	21	22	2
70	.	8	9	9	13	15	16	21	22	23	2,5
71,5	8	9	10	11	14	16	17	22	23	24	3
73	9	10	11	12	15	17	18	23	24	25	3,5
74,5	10	11	12	13	16	19	20	24	25	26	4
76	11	12	13	14	17	20	21	25	26	27	5
77,5	12	13	14	15	18	21	22	26	27	28	7
79	13	14	15	16	19	22	23	27	28	29	8
80,5	14	15	17	17	20	23	24	28	29	30	10
82	.	16	18	18	21	24	25	29	30	31	12
83,5	15	17	19	20	23	25	26	30	31	32	13
85	16	18	20	21	24	26	27	31	32	33	16
86,5	17	19	21	22	25	27	29	32	33	34	18
88	18	20	22	23	26	28	30	33	34	35	21
89,5	19	21	23	24	27	29	31	34	35	36	24
91	20	22	24	25	28	30	32	35	36	37	27
92,5	21	23	26	26	29	31	33	36	37	38	31
94	22	24	27	27	30	32	34	37	38	39	34
95,5	23	25	28	29	31	33	35	38	39	40	38
97	24	26	29	30	32	35	36	39	40	41	42
98,5	25	28	30	31	33	36	37	40	41	42	46
100	26	29	31	32	34	37	39	41	42	.	50
101,5	27	30	33	33	35	38	40	42	43	43	54
103	28	31	34	34	36	39	41	43	44	44	58
104,5	29	32	35	35	37	41	42	44	45	45	62
106	30	33	36	36	38	42	43	45	46	46	66
107,5	31	34	37	37	39	43	44	46	47	47	69
109	32	36	38	39	41	44	45	47	48	48	73
110,5	33	37	39	40	42	45	46	48	49	49	76
112	34	38	40	41	43	47	48	49	50	50	79
113,5	35	39	42	42	44	48	49	50	51	51	82
115	36	40	43	43	45	49	50	51	52	52	84
116,5	37	41	44	44	46	50	51	52	53	53	86
118	38	42	45	45	47	51	52	53	54	54	88
119,5	39	43	46	46	49	52	53	54	55	55	90
121	40	44	47	48	50	53	54	55	56	56	92
122,5	41	46	48	49	51	54	55	56	57	57	93
124	42	47	50	50	52	55	56	57	58	58	95
125,5	43	48	51	51	53	56	57	58	59	59	96
127	44	49	52	52	55	57	58	59	60	60	96,5
128,5	45	50	53	53	56	58	59	60			97
130	46	51	54	54	57	59	60				97,5
131,5	47	52	55	55	58	60					98
133	48	53	56	57	59						98,5
134,5	49	54	58	58	60						99
136	50	55	59	59							99,2
137,5	51	57	60	60							99,4
139	52	58									99,5
140,5	53	59									99,6
142	54	60									99,7
143,5	55										99,8
145	56										99,9

Fortsetzung Tabelle I a

IQ	15;0 bis 15;5	15;6 bis 15;11	16;0 bis 16;5	16;6 bis 16;11	17;0 bis 17;5	17;6 bis 17;11	18;0 bis 18;11	19+	PR
55	16	18	18	19	21	24	25	24	0,1
56,5	.	19	19	20	.	25	26	25	0,2
58	17	20	20	21	22	26	.	26	0,3
59,5	18	21	21	22	23	27	27	.	0,4
61	19	22	22	23	24	.	28	27	0,5
62,5	20	23	23	24	25	28	.	28	0,6
64	21	24	24	25	26	29	29	29	0,8
65,5	22	.	.	.	27	.	30	.	1
67	23	25	25	26	.	30	31	30	1,5
68,5	24	26	26	27	28	31	.	31	2
70	.	27	27	28	29	32	32	32	2,5
71,5	25	28	28	29	30	.	33	.	3
73	26	29	29	30	31	33	34	33	3,5
74,5	27	30	30	.	32	34	.	34	4
76	28	.	.	31	.	35	35	35	5
77,5	29	31	31	32	33	.	.	36	7
79	30	32	32	33	34	36	36	.	8
80,5	31	33	33	34	35	37	37	37	10
82	32	34	34	35	36	38	38	38	12
83,5	33	35	35	.	37	.	39	39	13
85	34	36	36	36	38	39	.	.	16
86,5	35	37	37	37	.	40	40	40	18
88	36	.	.	38	39	41	41	41	21
89,5	37	38	38	39	40	.	42	42	24
91	38	39	39	40	41	42	.	.	27
92,5	39	40	40	.	42	43	43	43	31
94	40	41	41	41	43	44	44	44	34
95,5	41	42	42	42	44	.	.	45	38
97	42	43	43	43	.	45	45	.	42
98,5	.	.	44	44	45	46	46	46	46
100	43	44	45	45	46	47	47	47	50
101,5	44	45	46	47	47	.	.	48	54
103	45	46	47	48	48	48	48	49	58
104,5	46	47	48	49	49	49	49	.	62
106	.	48	49	50	50	50	.	50	66
107,5	47	49	50	.	.	.	50	51	69
109	48	.	.	51	51	51	51	52	73
110,5	49	50	51	52	52	52	52	.	76
112	50	51	52	.	.	.	.	53	79
113,5	51	52	53	53	53	53	53	54	82
115	52	53	.	54	54	54	54	55	84
116,5	53	54	54	55	55	55	55	.	86
118	54	55	55	.	.	.	.	56	88
119,5	55	56	56	56	56	56	56	57	90
121	56	.	.	57	57	57	57	58	92
122,5	57	57	57	58	58	58	58	.	93
124	58	58	58	.	.	.	.	59	95
125,5	59	59	59	59	59	59	59	60	96
127	60	60	60	60	60	60	60	.	96,5
128,5	.	.	.	.	.	.	.	.	97
130	.	.	.	.	.	.	.	.	97,5

Tabelle 1b Schulnormen zum Progressiven Matrizen-Test für *taube* Kinder und Jugendliche

IQ	Grund- und Hauptschulklassen (PMT-Rohpunkte)									PR
	1	2	3	4	5	6	7	8	9	
55					1	3	5	7	19	0,1
56,5					2	4	6	8	20	0,2
58				1	3	5	7	9	21	0,3
59,5	1	1	1	2	4	6	8	10	22	0,4
61	2	2	2	3	5	7	9	11	23	0,5
62,5	3	3	3	4	6	8	10	12	.	0,6
64	4	4	.	5	7	9	11	13	24	0,8
65,5	.	.	4	6	8	10	12	14	25	1
67	5	5	5	7	9	11	13	15	26	1,5
68,5	.	.	6	8	10	12	14	16	.	2
70	6	6	.	.	11	13	15	17	27	2,5
71,5	.	.	7	9	12	14	16	18	28	3
73	7	7	8	10	13	15	17	19	29	3,5
74,5	.	.	9	11	14	16	18	20	.	4
76	8	8	.	12	15	17	20	21	30	5
77,5	.	9	10	13	16	18	21	22	31	7
79	9	10	11	14	17	19	22	23	32	8
80,5	.	.	12	15	18	20	23	24	33	10
82	10	11	13	16	19	21	24	25	.	12
83,5	.	.	.	17	.	23	25	26	34	13
85	11	12	14	18	20	24	26	27	35	16
86,5	.	.	15	.	21	25	27	28	36	18
88	12	13	16	19	22	26	28	29	.	21
89,5	.	.	.	20	23	27	29	30	37	24
91	13	14	17	21	24	28	30	31	38	27
92,5	.	15	18	22	25	29	31	32	39	31
94	14	.	19	23	26	30	32	33	40	34
95,5	.	16	.	24	27	32	33	34	.	38
97	.	.	20	25	28	33	34	35	41	42
98,5	15	17	21	26	29	34	35	36	42	46
100	.	.	22	27	30	35	36	37	43	50
101,5	.	18	.	28	31	36	37	38	.	54
103	16	.	23	29	32	37	38	39	44	58
104,5	.	19	24	30	33	38	39	40	45	62
106	.	20	25	31	34	39	40	41	46	66
107,5	17	.	.	32	35	40	41	42	47	69
109	.	21	26	33	36	.	42	43	.	73
110,5	.	.	27	34	37	41	43	44	48	76
112	18	22	28	35	38	42	44	45	49	79
113,5	.	.	.	36	39	43	45	46	50	82
115	19	23	29	37	40	44	46	47	.	84
116,5	.	24	30	38	41	45	47	48	51	86
118	20	.	31	39	42	46	48	49	52	88
119,5	.	25	.	40	43	47	49	50	53	90
121	.	.	32	41	44	48	50	51	54	92
122,5	21	26	33	42	45	49	51	52	.	93
124	.	.	34	44	46	50	52	53	55	95
125,5	22	27	.	45	47	.	53	54	56	96
127	.	.	35	46	48	51	54	55	57	96,5
128,5	23	28	36	47	49	52	55	56	.	97
130	.	29	37	48	50	53	56	57	58	97,5
131,5	24	.	.	49	51	54	57	58	59	98
133	.	30	38	50	53	55	58	59	60	98,5
134,5	.	.	39	51	54	56	.	60	.	99
136	25	31	40	52	55	57	60	.	.	99,2
137,5	.	.	.	53	56	58	.	.	.	99,4
139	.	32	41	54	57	59	.	.	.	99,5
140,5	26	33	42	55	58	60	.	.	.	99,6
142	.	.	43	56	59	.	.	.	.	99,7
143,5	.	34	.	57	60	.	.	.	.	99,8
145	27	35	44	58	.	.	.	.	.	99,9

Tabelle I c Schulnormen zum Progressiven Matrizen-Test für *resthörige/schwerhörige* Kinder und Jugendliche

IQ	Grund- und Hauptschulklassen (PMT-Rohpunkte)									PR
	1	2	3	4	5	6	7	8	9	
55						2	4	5	20	0,1
56,5						3	5	6	.	0,2
58						4	6	7	21	0,3
59,5					1	5	7	8	22	0,4
61					.	6	8	9	23	0,5
62,5					2	7	9	10	24	0,6
64				1	3	8	10	11	25	0,8
65,5	1	1	1	2	4	9	11	12	.	1
67	.	.	.	3	5	10	12	13	26	1,5
68,5	2	2	2	.	6	11	13	14	27	2
70	.	3	3	4	7	12	14	15	28	2,5
71,5	3	.	4	5	8	13	15	16	29	3
73	.	4	5	6	9	14	16	17	30	3,5
74,5	4	5	6	7	10	15	17	18	.	4
76	5	6	.	8	11	16	18	19	31	5
77,5	.	.	7	9	12	17	19	20	32	7
79	6	7	8	.	13	18	20	21	33	8
80,5	7	8	9	10	14	19	21	22	34	10
82	8	.	10	11	15	20	22	23	35	12
83,5	.	9	11	12	16	21	23	24	.	13
85	9	10	12	13	17	22	24	25	36	16
86,5	10	11	13	14	18	23	25	26	37	18
88	.	12	14	15	19	24	26	27	38	21
89,5	11	13	15	16	20	25	27	28	39	24
91	.	.	16	17	21	26	28	29	40	27
92,5	12	14	17	18	22	27	29	30	.	31
94	.	15	18	19	23	29	30	31	41	34
95,5	13	16	19	.	24	30	31	32	42	38
97	.	17	20	20	25	31	32	33	43	42
98,5	14	18	21	21	26	32	33	34	44	46
100	.	.	.	22	27	33	34	35	45	50
101,5	15	19	22	23	28	34	35	36	.	54
103	.	20	23	24	29	35	36	37	46	58
104,5	.	.	.	25	30	36	37	38	47	62
106	16	21	24	.	31	37	38	39	48	66
107,5	.	22	25	26	32	38	39	40	.	69
109	17	.	.	27	34	39	40	42	49	73
110,5	.	23	26	28	35	40	41	43	50	76
112	.	24	27	29	36	42	42	44	51	79
113,5	18	.	.	30	37	43	43	45	.	82
115	19	25	28	31	38	44	44	46	52	84
116,5	.	26	29	32	39	45	45	47	53	86
118	.	.	.	33	41	46	47	48	54	88
119,5	20	27	30	34	42	47	48	50	.	90
121	.	28	31	35	43	48	49	51	55	92
122,5	.	.	.	36	44	49	50	52	56	93
124	21	29	32	37	45	50	51	53	57	95
125,5	.	30	.	38	46	51	52	54	.	96
127	22	.	33	39	47	52	53	55	58	96,5
128,5	.	31	34	40	49	53	54	56	59	97
130	23	32	.	41	50	55	55	58	.	97,5
131,5	.	.	35	42	51	56	56	59	60	98
133	.	33	36	43	52	57	57	60	.	98,5
134,5	24	34	.	44	53	58	58	.	.	99
136	.	.	37	45	54	59	59	.	.	99,2
137,5	.	35	38	46	56	60	60	.	.	99,4
139	25	36	.	47	57	.	.	.	.	99,5
140,5	.	.	39	48	58	.	.	.	.	99,6
142	.	37	40	49	59	.	.	.	.	99,7
143,5	26	38	.	50	60	.	.	.	.	99,8
145	.	39	41	51	.	.	.	.	.	99,9

Tabelle I d Schulnormen zum Progressiven Matrizen-Test für Hörgeschädigte weiterführender gehobener Bildungseinrichtungen (R = Realschule und W = Wirtschaftsschule)

IQ	5 R*	6 R	7 R	8 R	9 R	10 R	10 W	11 W	12 W	PR
55	17	24	25	30	34	44	23	35	41	0,1
56,5	18	.	26	31	35	.	24	.	.	0,2
58	19	25	27	.	.	.	25	36	42	0,3
59,5	20	26	.	32	36	45	26	.	.	0,4
61	.	.	28	33	.	.	.	37	.	0,5
62,5	21	27	.	.	37	.	27	.	43	0,6
64	22	.	29	34	.	46	28	38	.	0,8
65,5	.	28	.	35	38	.	29	.	.	1
67	23	29	30	.	.	.	30	39	44	1,5
68,5	24	.	31	36	39	47	.	.	.	2
70	25	30	.	37	.	.	31	40	45	2,5
71,5	.	31	32	.	40	.	32	.	.	3
73	26	.	.	38	.	.	33	41	.	3,5
74,5	27	32	33	.	41	48	.	.	46	4
76	28	.	34	39	.	.	34	42	.	5
77,5	.	33	.	40	42	.	35	.	.	7
79	29	34	35	.	.	49	36	.	47	8
80,5	30	.	.	41	43	.	37	43	.	10
82	31	35	36	.	.	.	.	.	.	12
83,5	.	36	.	42	44	50	38	45	48	13
85	32	.	37	43	.	.	39	.	.	16
86,5	33	37	38	.	45	.	40	.	.	18
88	.	.	.	44	.	51	41	46	49	21
89,5	34	38	39	45	46	.	.	.	.	24
91	35	39	.	.	.	.	42	.	50	27
92,5	36	.	40	46	47	52	43	47	.	31
94	.	40	41	.	.	.	44	.	.	34
95,5	37	41	.	47	48	.	45	.	51	38
97	38	.	42	48	.	53	.	48	.	42
98,5	39	42	.	.	49	.	46	.	.	46
100	.	.	43	49	.	.	47	49	52	50
101,5	40	43	.	.	50	54	48	.	.	54
103	41	44	44	50	.	.	.	50	.	58
104,5	.	.	45	.	51	.	49	.	53	62
106	42	45	.	51	55	50	.	.	.	66
107,5	43	46	46	.	52	.	.	51	.	69
109	44	.	.	52	.	.	.	.	54	73
110,5	.	47	47	.	53	56	51	52	.	76
112	45	.	48	53	.	.	52	.	55	79
113,5	46	48	.	54	54	.	.	53	.	82
115	47	49	49	54	.	57	53	.	.	84
116,5	.	.	50	.	55	.	.	54	56	86
118	48	50	.	55	.	.	.	.	.	88
119,5	49	51	51	.	56	58	54	55	.	90
121	.	.	52	56	.	.	55	.	57	92
122,5	50	52	.	57	57	.	.	56	.	93
124	51	.	53	57	.	59	56	.	.	95
125,5	52	53	54	.	58	.	.	57	58	96
127	.	54	.	58	.	.	57	.	.	96,5
128,5	53	.	55	.	59	60	.	58	.	97
130	54	55	56	59	.	.	58	.	59	97,5
131,5	55	56	.	60	60	.	.	59	.	98
133	.	.	57	.	.	.	59	.	60	98,5
134,5	56	57	.	.	.	.	.	.	.	99
136	57	.	58	.	.	.	60	60	.	99,2
137,5	.	58	59	.	.	.	.	.	.	99,4
139	58	59	.	.	.	.	.	.	.	99,5
140,5	59	.	60	.	.	.	.	.	.	99,6
142	60	60	.	.	.	.	.	.	.	99,7
143,5	.	.	.	.	.	.	.	.	.	99,8
145	.	.	.	.	.	.	.	.	.	99,9

\* Die R- und W-Klassen sind nach Schuljahren fortlaufend durchnummeriert.

Tabelle I e Berufsschulnormen zum Progressiven Matrizen-Test für Hörgeschädigte

IQ	Berufsschüler (insgesamt)	Textilklass u. Anlernlinge	Technisches Handwerk	Tech. Zeichner u. Zahntechniker	PR
55	29	25	27	34	0,1
56,5	30	.	28	.	0,2
58	.	26	.	35	0,3
59,5	31	27	29	.	0,4
61	32	.	.	36	0,5
62,5	.	28	30	.	0,6
64	33	29	.	37	0,8
65,5	.	.	31	.	1
67	34	30	32	38	1,5
68,5	.	.	.	.	2
70	35	31	33	39	2,5
71,5	.	32	.	.	3
73	36	.	34	40	3,5
74,5	37	33	.	.	4
76	.	34	35	.	5
77,5	38	.	.	41	7
79	.	35	36	.	8
80,5	39	.	37	42	10
82	.	36	.	43	12
83,5	40	37	38	43	13
85	.	.	.	.	16
86,5	41	38	39	44	18
88	42	39	.	.	21
89,5	.	.	40	45	24
91	43	40	41	.	27
92,5	.	.	.	46	31
94	44	41	42	.	34
95,5	.	42	.	47	38
97	45	.	43	.	42
98,5	.	43	.	48	46
100	46	44	44	.	50
101,5	47	.	45	49	54
103	.	45	.	.	58
104,5	48	.	46	50	62
106	.	46	.	.	66
107,5	49	47	47	51	69
109	.	.	.	.	73
110,5	50	48	48	52	76
112	.	49	49	.	79
113,5	51	.	.	53	82
115	52	50	50	.	84
116,5	.	.	.	54	86
118	53	51	51	.	88
119,5	.	52	52	55	90
121	54	.	.	.	92
122,5	.	53	53	56	93
124	55	54	54	.	95
125,5	.	.	.	57	96
127	56	55	55	.	96,5
128,5	.	56	.	58	97
130	57	.	56	.	97,5
131,5	.	57	57	.	98
133	58	.	.	59	98,5
134,5	.	58	58	.	99
136	59	59	.	60	99,2
137,5	.	.	59	.	99,4
139	60	60	.	.	99,5
140,5	.	.	60	.	99,6
142	.	.	.	.	99,7
143,5	.	.	.	.	99,8
145	.	.	.	.	99,9

Tabelle II a Alters-Prozentränge (PR) zum Benton-Test (Zeichenform C und Wahlform G) für taube/resthörige Schüler

PR (cum f %)	Altersgruppen											
	7;6 – 10;5 (N = 39)				10;6 – 13;5 (N = 51)				13;6 – 16;5 (N = 34)			
	RP	C	F		RP	C	F		RP	C	F	
0,5												
1												
1,5												
2					0		10		2		13	
2,5	0		10		3		12					
3												1
3,5												14
4					2		8		3		12	
5					3		7					
7	1		9						4		6	
8	2		8									
10				4	4		6					9
12												6
13				5					9		6	
16				6					10		5	
18	3		7	7	5		5		5		5	
21				8								
24												11
27	4		6									4
31												
34				8				11	4			
38				9						6	4	
42					6		4					12
46												3
50				10			5					
54	5		5					12	3			13
58												2
62				11			4					
66					7		3			7	3	
69												
73				12			3		13		2	
76	6		4									
79					8		2					14
82												1
84								14	1			
86										8	2	
88												
90												
92	7		3									
93												
95				13			2					
96												
96,5												
97										9	1	
97,5												
98					9		1					
98,5												
99												
99,5												
100	8		2	14	1	10	0	15	0	10	0	15
												0

Abk.: C = Zeichenform C, G = Wahlform G; RP = Rohpunkte (Anzahl der richtigen Reproduktionen bzw. Wahltreffer), F = Fehler



**Tabelle II b** Altersnormen zum Benton-Test (Zeichenform C und Wahlform G) sowie Fehlerkategorien der Zeichenform C (%-Anteile) für taube/resthörige Schüler

Benton-Test	Zeichenform C			Wahlform G		
	Altersgruppen			Altersgruppen		
	7;6-10;5 (N = 39)	10;6-13;5 (N = 51)	13;6-16;5 (N = 34)	7;6-10;5 (N = 39)	10;6-13;5 (N = 51)	13;6-16;5 (N = 34)
<b>I. Rohpunkte</b> (Anzahl der richtigen Reproduktionen bzw. Treffer)						
Arithm. Mittel	5,1	6,9	7,0	9,9	11,9	12,6
Stand. Abweichg. (Sigma-Wert)	1,9	2,0	1,5	3,1	2,5	2,7
Max.-Wert	10,0	10,0	10,0	14,0	15,0	15,0
Q <sub>3</sub> = 75. PR	5,8	7,6	7,6	12,5	13,5	13,8
Med = 50. PR	4,9	6,6	6,7	10,5	11,8	12,7
Q <sub>1</sub> = 25. PR	3,8	5,6	5,7	7,7	10,8	11,5
Min.-Wert	0,0	0,0	4,0	3,0	2,0	1,0
<b>II. Fehler (abs.)</b> (Anzahl der Fehler)						
Arithm. Mittel	4,9	3,3	3,1	5,1	3,1	2,5
Stand. Abweichg. (Sigma-Wert)	1,9	2,4	1,5	3,0	2,8	2,4
Max.-Wert	10,0	10,0	6,0	12,0	13,0	14,0
Q <sub>3</sub> = 75. PR	6,3	4,4	4,3	7,3	4,4	3,5
Med = 50. PR	5,1	3,4	3,3	4,5	3,2	2,3
Q <sub>1</sub> = 25. PR	4,2	2,4	2,4	2,5	1,5	1,2
Min.-Wert	0,0	0,0	0,0	1,0	0,0	0,0
<b>III. Fehlerkategorien</b> (Anteile in %)						
A = Auslassungen	12,1	11,2	4,2			
E = Entstellungen	54,0	54,2	51,7			
P = Perseverationen	10,1	8,9	17,5			
D = Drehungen	10,7	16,1	17,5			
F = Fehlplatzierungen	10,0	8,8	7,0			
G = Größenfehler	3,1	0,8	2,1			

*Anm.:* Bei der diagnostischen Interpretation vorstehender Normdaten ist zu beachten, daß bezüglich der Fehlerzahl bzw. des Fehler-% die Min.-Werte die beste und die Max.-Werte die schlechteste Leistung repräsentieren; analog markiert hier der Q<sub>1</sub>-Wert die obere Grenze, der Q<sub>3</sub>-Wert die untere Grenze des mittleren Bereiches, in dem 50% aller Fälle liegen.

Tabelle III Altersnormen zum HAWIK (Handlungsteil) für taube/resthörige Schüler

Wertpunkttable zu den HAWIK-Handlungstests Umwandlung der Rohpunkte (RP) in Wertpunkte (WP)						IQ-Transformation <sup>2</sup> der Gesamt-Wertpunkte (WP-Summe aus 5 Subtests) im HAWIK-Handlungsteil			
Standard- wertskala <sup>1</sup>	Zahlen- Symbol- Test	Bilder- ergän- zen	Bilder- ordnen	Mosaik- Test	Figuren- legen	WP-Summe <sup>3</sup>		IQ	
WP	RP	RP	RP	RP	RP				
Altersgruppe 8;0-9;11	0	0- 7	0-1	—	—	0- 3	12-13	3- 4	55
	1	8- 9	2	—	—	4- 5	14-15	5- 6	57
	2	10-11	3	0	—	6	16-17	7- 8	59
	3	12-13	—	1- 2	—	7- 8	18	9-10	61
	4	14-15	4	3- 4	0	9	19-20	11-12	63
	5	16	5	5- 6	1- 4	10-11	21-22	13-14	65
	6	17-18	—	7- 8	5- 7	12-13	23	15-16	67
	7	19-20	6	9-10	8-11	14	24-25	17-18	69
	8	21-22	—	11-12	12-14	15-16	26-27	19-20	71
	9	23-24	7	13-14	15-18	17	28	21-22	73
	10	25-26	8	15-17	19-21	18-19	29-30	23-24	75
	11	27-28	—	18-19	22-25	20	31-32	25-26	77
	12	29-30	9	20-21	26-28	21-22	33	27-28	79
	13	31-32	10	22-23	29-32	23-24	34-35	29-30	81
	14	33-34	—	24-25	33-35	25	36	31-32	83
	15	35	11	26-27	36-39	26-27	37-38	33-34	85
	16	36-37	12	28-29	40-42	28	39	35-36	87
	17	38-39	—	30-31	43-46	29-30	40-41	37-38	89
	18	40-41	13	32-33	47-49	31-32	42-43	39-40	91
	19	42-43	14	34-36	50-53	33	44-45	41-42	93
20	44+	15+	37+	54-55	34	46	43-44	95	
Altersgruppe 10;0-11;11	0	0-24	0	0	—	0- 3	47-48	45-46	97
	1	25	1	1- 2	—	4- 5	49	47-48	99
	2	—	2	3- 5	—	6- 7	50-51	49-50	101
	3	26	3	6- 7	0	8	52	51-52	103
	4	—	4	8- 9	1- 2	9-10	53-54	53-54	105
	5	27	5	10-12	3- 6	11-12	55-56	55-56	107
	6	28	6	13-14	7-10	13	57	57-58	109
	7	—	7	15-16	11-14	14-15	58-59	59-60	111
	8	29	8	17-19	15-18	16-17	60-61	61-62	113
	9	—	9	20-21	19-22	18	62	63-64	115
	10	30	10	22-23	23-27	19-20	63-64	65-66	117
	11	—	11	24-25	28-31	21-22	65-66	67-68	119
	12	31	12	26-28	32-35	23	67	69-70	121
	13	32	13	29-30	36-39	24-25	68-69	71-72	123
	14	—	14	31-32	40-43	26	70-71	73-74	125
	15	33	15	33-35	44-47	27-28	72	75-76	127
	16	—	16	36-37	48-51	29-30	73-74	77-78	129
	17	34	17	38-39	52-55	31	75-76	79-80	131
	18	35	18	40-41	—	32-33	77	81-82	133
	19	—	19	42-44	—	34	78-79	83-84	135
20	36+	20	45+	—	—	80-81	85-86	137	
						82	87-88	139	
						83-84	89-90	141	
						85-86	91-92	143	
						87-88	93-94	145	

<sup>1</sup> Die Standard(Wertpunkt)-Skala ist definiert durch  $M = 10$  und  $s = 3$ ; siehe auch Tabelle VIII im Anhang.

<sup>2</sup> IQ = Abweichungs-IQ ( $M = 100$ ,  $s = 15$ ); siehe auch Tabelle VIII im Anhang.

<sup>3</sup> Altersgruppe 8;0 bis 9;11 = linke Spalte; Altersgruppe 10;0 bis 11;11 = rechte Spalte.

Wertpunkttabelle zu den HAWIK-Handlungstests Umwandlung der Rohpunkte (RP) in Wertpunkte (WP)						IQ-Transformation <sup>2</sup> der Gesamt-Wertpunkte (WP-Summe aus 5 Subtests) im HAWIK-Handlungsteil			
Standard- wertskala <sup>1</sup>	Zahlen- Symbol- Test	Bilder- ergän- zen	Bilder- ordnen	Mosaik- Test	Figuren legen	WP-Summe <sup>3</sup>		IQ	
WP	RP	RP	RP	RP	RP				
Altersgruppe 12;0-12;11	0	0- 3	0- 3	0- 7	—	0- 5	14	9	55
	1	4- 7	4	8- 9	—	6- 7	15-16	10-11	57
	2	8-10	5	10	0	8	17-18	12-13	59
	3	11-14	—	11-12	1- 4	9-10	19	14	61
	4	15-17	6	13-14	5- 8	11	20-21	15-16	63
	5	18-21	7	15	9-12	12-13	22	17-19	65
	6	22-24	8	16-17	13-16	14-15	23-24	20	67
	7	25-28	—	18-19	17-20	16	25	21-22	69
	8	29-31	9	20	21-24	17-18	26-27	23-24	71
	9	32-35	10	21-22	25-28	19	28	25	73
	10	36-38	11	23-24	29-32	20-21	29-30	26-27	75
	11	39-42	12	25	33-36	22-23	31-32	28-29	77
	12	43-45	—	26-27	37-40	24	33	30	79
	13	46-49	13	28-29	41-44	25-26	34-35	31-32	81
	14	50-52	14	30	45-48	27	36	33-34	83
	15	53-56	15	31-32	49-52	28-29	37	35-36	85
	16	57-59	—	33-34	53-55	30	38-39	37-38	87
	17	60-63	16	35	—	31-32	40-41	39-40	89
	18	64-66	17	36-37	—	33-34	42	41	91
	19	67-70	18	38-39	—	—	43	42-43	93
20	71+	19-20	40+	—	—	44-45	44-45	95	
Altersgruppe 13;0-13;11	0	0- 7	0- 5	—	—	0-13	46-47	46-47	97
	1	8-10	6	—	0	14-15	48	48-49	99
	2	11-14	7	—	1- 3	16	49-50	50-51	101
	3	15-18	—	0	4- 7	17	51	52	103
	4	19-21	8	1- 2	8-11	18	52-53	53-54	105
	5	22-25	9	3- 6	12-15	19	54-55	55-56	107
	6	26-29	10	7-11	16-19	20	56	57	109
	7	30-33	11	12-15	20-23	21	57	58-59	111
	8	34-36	—	16-20	24-28	22	58-59	60-61	113
	9	37-40	12	21-24	29-32	23	60	62-63	115
	10	41-44	13	25-28	33-36	24	61-62	64-65	117
	11	45-47	14	29-33	37-40	25	63-64	66-67	119
	12	48-51	15	34-37	41-44	26	65	68	121
	13	52-55	16	38-42	45-48	27	66-67	69-70	123
	14	56-59	—	43-46	49-52	28	68-69	71-72	125
	15	60-62	17	47-51	53-55	29	70	73	127
	16	63-66	18	52-55	—	30	71-72	74-75	129
	17	67-70	19	56-57	—	31	73	76-77	131
	18	71-73	20	—	—	32	74	78-79	133
	19	74-77	—	—	—	33	75-76	80-81	135
20	78+	—	—	—	34	77-78	82-83	137	
						79	84	139	
						80-81	85-86	141	
						82	87-88	143	
						83-84	89-90	145	

<sup>1</sup> Die Standard(Wertpunkt)-Skala ist definiert durch  $M = 10$  und  $s = 3$ ; siehe auch Tabelle VIII im Anhang.

<sup>2</sup> IQ = Abweichungs-IQ ( $M = 100$ ,  $s = 15$ ); siehe auch Tabelle VIII im Anhang.

<sup>3</sup> Altersgruppe 12;0 bis 12;11 = linke Spalte; Altersgruppe 13;0 bis 13;11 = rechte Spalte.

Fortsetzung Tabelle III

Wertpunkttabelle zu den HAWIK-Handlungstests Umwandlung der Rohpunkte (RP) in Wertpunkte (WP)						IQ-Transformation <sup>2</sup> der Gesamt-Wertpunkte (WP-Summe aus 5 Subtests) im HAWIK-Handlungsteil		
Standard- wertskala <sup>1</sup>	Zahlen- Symbol- Test	Bilder- ergän- zen	Bilder- ordnen	Mosaik- Test	Figuren- legen	WP-Summe	IQ	
WP	RP	RP	RP	RP	RP			
Altersgruppe 14;0-15;11	0	0-17	0-5	0- 7	0- 4	0-15	13	55
	1	18-20	6	8- 9	5- 7	16	14-15	57
	2	21-23	7	10-11	8-11	17	16-17	59
	3	24-26	—	12-13	12-14	18	18	61
	4	27-29	8	14-15	15-18	19	19-20	63
	5	30-32	9	16-17	19-21	20	21	65
	6	33-35	10	18-19	22-24	21	22-23	67
	7	36-38	11	20-21	25-28	22	24-25	69
	8	39-42	12	22-23	29-31	23	26	71
	9	43-45	—	24-25	32-34	24	27	73
	10	46-48	13	26-27	35-38	25	28-29	75
	11	49-51	14	28-29	39-41	26	30-31	77
	12	52-54	15	30-31	42-44	27	32	79
	13	55-57	16	32-33	45-48	—	33-35	81
	14	58-60	17	34-35	49-51	28	36	83
	15	61-63	—	36-37	52-55	29	37	85
	16	64-66	18	38-39		30	38-39	87
	17	67-69	19	40-41		31	40-41	89
	18	70-72	20	42-45		32	42	91
	19	73-75		44-45		33	43-44	93
	20	76+		46+		34	45	95
							46-47	97
							48-49	99
							50	101
							51	103
							52-53	105
							54-55	107
							56	109
							57-59	111
							60	113
							61	115
							62-63	117
							64-65	119
							66	121
							67-68	123
							69	125
							70-71	127
							72-73	129
							74	131
							75	133
							76-77	135
							78-79	137
							80	139
							81-83	141
							84	143
							85-86	145

<sup>1</sup> Die Standard(Wertpunkt)-Skala ist definiert durch  $M = 10$  und  $s = 3$ ; siehe auch Tabelle VIII im Anhang.

<sup>2</sup> IQ = Abweichungs-IQ ( $M = 100$ ,  $s = 15$ ); siehe auch Tabelle VIII im Anhang.

Tabelle IV a Schul-Prozentränge (PR) im Frankfurter Wortschatztest (WST Form I bzw. II) für taube/resthörige Hauptschüler<sup>1</sup> unter Verwendung des WST als Powertest

PR (cum f %)	Hauptschulklassen (WST-Rohpunkte)									
	4	5	6	7	8	9	10			
	I <sup>2</sup> N=23	I N=48	I N=52	I II N=60 N=19	I II N=52 N=27	I II N=38 N=19	I II N=22 N=4 <sup>3</sup>			
0,5										
1										
1,5										
2		1			1					
2,5						1				
3										
3,5				2	3					
4	5									
5			2	3		2	2			
7					5					
8				3		4				3
10			3		2	4	4			
12		2								
13	6			4	3	5				6
16			4	4	7	5	5			
18			5		4					
21		3		5	5					
24			6		5					2
27		4			8					
31		5		6	6	6	6			7
34	7	6		7			8			8
38			7	7	9	7				
42	8			8	7	9	9			9
46			8	8		8				
50	9	7		9	8	10	10			11
54			9	9	9					12
58		8		10	11	9				13
62							11			
66	10		10	10	10					
69		9				10				14
73				11	11	12				
76		10	11	11	12					15
79			12	12	12	12	12			12
82	11					13				
84		11		13						
86				12	13	11				18
88				13	14		13			
90	12		14	14						20
92		12	15			15				
93				14						
95		13	16	16	15	16	17			
96	13		17			16				
96,5				17		15				
97										
97,5							17			
98		14	18	18	18					
98,5										
99										
99,5										
100	19	23	20	22	17	30	18			24
										16

<sup>1</sup> Versuchsweise wurde auch ein kleines Sample der 4. (Grundschul-)Klasse mit dem für die 5. bzw. 6. Klasse (Hörender) zugeschnittenen WST I untersucht. Siehe auch die Erläuterungen auf Seite 202f. in diesem Buch. – <sup>2</sup> WST-Form (I = 5-6, II = 7-8) und Stichproben-N. – <sup>3</sup> Mit Rücksicht auf die geringe Pbn-Zahl kann diesen Daten vorerst nur grobe Orientierungsfunktion zuerkannt werden.

**Tabelle IV b Schul-Prozentränge (PR) im Frankfurter Wortschatztest (WST Form I bzw. II) für schwerhörige Hauptschüler unter Verwendung des WST als Powertest**

PR (cum f %)	5 <sup>I</sup> I <sup>2</sup> N = 23	6 I N = 36	7 I N = 20	II N = 22	8 I N = 28	II N = 19	9 I N = 19	II N = 13
0,5 1 1,5 2 2,5								
3 3,5 4 5 7 8 10 12 13 16	5       6	1     4	3  4		4  5  6		2  6  9	   3
18 21 24 27 31 34 38 42 46 50	7   8   10 11	5 6 7 8 9	5 6 7 8 9 10	2  4  5	7 8 9 10 11 12 7	1  3  6 7	10  11 12 15	5  8 10
54 58 62 66 69 73 76 79 82 84	12 13 14 15 16 18 19	10 11 12 13	11 12 13 14 15 16	6 7  8 9 10	13 14 15 17	9 11 12	16 17 18 19	 11 13 14
86 88 90 92 93 95 96 96,5 97 97,5	23 24  29	14 15 16 17 20	18 22	11 14	20 21 23 26	13 14 15	23 24	15
98 98,5 99 99,5 100	   30	   26	   26	  15	  29	  18	  30	  18

<sup>1</sup> In dieser 1. Hauptschulklasse sind praktisch noch alle Begabungen vertreten, da der Übertritt auf Realschule und/oder Gymnasium bei Hörgeschädigten in der Regel erst zum Ende des 5. Schuljahrs erfolgt. – 2 WST-Form (I = 5-6, II = 7-8) und Stichproben-Größe (N)

Tabelle IV c Schul-Prozentränge (PR) im Frankfurter Wortschatztest (WST Form I bzw. II) für hörgeschädigte Realschüler und Wirtschaftsschüler – Testbearbeitungszeit: 20 Min.

PR (cum f %)	Realschulklassen 5-10 (WST-Rohpunkte)								
	5 <sup>1</sup> 12 N=21	6 I N=10	7 I N=11	II N=19	8 I N=23	II N=18	9 I N=21	10 I N=20	II N=20
3									
3,5									
4					13				
5	8			5		8	18	18	18
7									
8					15				
10		14	15			13	21	21	21
12									
13	9				16		24	22	
16				7					22
18			20				27	27	
21		19			19				24
24	10						28	32	27
27	11		23	8	22	15	30		
31		21		9	23			39	32
34						20	32	40	
38	12		29		28		34		38
42	14	33		10	31		35		
46			30	12	33	22	37	43	39
50		39					39	52	
54			31	15	35	32			40
58	18			17			42		
62	21	42	34	18	39		45	53	
66	22					34	47		43
69					46		48	55	
73			39	20		35			
76					49	36	50	57	52
79		43		21					
82	25		41				53	58	
84						38			53
86	30				51		55		
88						39			
90	33	56	51	30			56	59	55
92									
93									
95	37			33		49	57		58
96					52				
96,5									
97									
97,5									
98									
98,5									
99									
99,5									
100	39	58	52	47	56	54	58	60	59

<sup>1</sup> Die Klassen sind jeweils nach Schuljahren durchnummeriert. Von der 9. Realschulklasse standen keine WST-Unterlagen der Form II (7-8) zur Verfügung.

<sup>2</sup> WST-Form (I = 5-6, II = 7-8) und Stichproben-Größe (N).

Fortsetzung Tabelle IV c

PR (cum f %)	Wirtschaftsschulklassen 10-12 (WST-Rohpunkte)					
	10 I N=9	II N=9	11 I N=9	II N=9	12 I N=11	II N=10
3						
3,6						
4						
5						
7						
8						
10		2	8	6	7	8
12						
13						
16						
18					12	
21	5	5	13	7		9
24						
27					14	
31						21
34	10	6	20	17		
38					24	
42	16			18		33
46		9	29			
50						36
54	19	10	31	23	45	
58						
62					46	42
66	23	17	50	45		
69						43
73					48	
76						
79	37	32	53	50		44
82					53	
84						
86						
88	51	41	54	51		
90					54	55
92						
93						
95						
96						
96,5						
97						
97,5						
98						
98,5						
99						
99,5						
100	55	42	57	54	60	60



Tabelle V a Altersnormen zum Bourdon-Figuren-Durchstreich-Test für taube Schüler  
(nach E. Martin 1961)

Bourdon-Figuren-Test	Altersgruppen			
	10;0-11;11 (N = 17)	12;0-13;11 (N = 16)	14;0-15;11 (N = 19)	16;0-17;11 (N = 5)
<b>I. Leistungsmenge</b> (Anzahl der durchgestrichenen Figuren)				
Arithm. Mittel	1042	1157	1252	1280
Max.-Wert	1285	1633	1660	1644
Q <sub>3</sub> = 75. PR	1198	1254	1396	1583
Med = 50. PR	1082	1165	1239	1251
Q <sub>1</sub> = 25. PR	884	1054	1151	988
Min.-Wert	588	771	837	925
<b>II. Leistungsgüte</b> (Arbeitssorgfalt)				
Fehler in %				
Arithm. Mittel	2,7	2,7	4,1	5,6
Max.-Wert	7,5	6,1	15,5	15,2
Q <sub>3</sub> = 75. PR	4,2	3,8	6,3	13,3
Med = 50. PR	2,1	2,3	3,1	2,5
Q <sub>1</sub> = 25. PR	0,7	1,2	1,6	1,2
Min.-Wert	0,4	0,7	0,7	1,1

Tabelle V b Altersnormen zum Bourdon-Figuren-Durchstreich-Test für resthörige Schüler  
(nach E. Martin 1961)

Bourdon-Figuren-Test	Altersgruppen			
	10;0-11;11 (N = 13)	12;0-13;11 (N = 14)	14;0-15;11 (N = 9)	16;0-17;11 (N = 6)
<b>I. Leistungsmenge</b> (Anzahl der durchgestrichenen Figuren)				
Arithm. Mittel	959	1091	1138	1237
Max.-Wert	1450	1450	1488	1664
Q <sub>3</sub> = 75. PR	1150	1281	1431	1557
Med = 50. PR	996	1101	1164	1201
Q <sub>1</sub> = 25. PR	780	867	800	975
Min.-Wert	179	704	700	825
<b>II. Leistungsgüte</b> (Arbeitssorgfalt)				
Fehler in %				
Arithm. Mittel	6,3	3,7	3,5	4,4
Max.-Wert	22,7	8,7	8,0	6,2
Q <sub>3</sub> = 75. PR	8,0	6,5	5,0	5,7
Med = 50. PR	6,2	3,0	3,4	5,0
Q <sub>1</sub> = 25. PR	1,5	1,1	1,8	1,3
Min.-Wert	1,1	0,4	1,5	1,1

Anm. zu Tab. Va und Vb: Bei der diagnostischen Interpretation vorstehender Normdaten ist zu beachten, daß bezüglich des Fehler-% die Min.-Werte die beste und die Max.-Werte die schlechteste Leistung repräsentieren; analog markiert hier (im Fehler-%) der Q<sub>1</sub>-Wert die obere Grenze, der Q<sub>3</sub>-Wert die untere Grenze des mittleren Bereiches, in dem 50% aller Fälle liegen.

Tabelle VI a Altersnormen zum Konzentrations-Verlaufs-Test für taube Schüler  
(KVT<sub>Taube</sub>)

K-V-T <sub>Taube</sub>	Altersgruppen			
	10;0-11;11 (N = 17)	12;0-13;11 (N = 19)	14;0-15;11 (N = 20)	16;0-17;11 (N = 6)
<b>I. Bearbeitungszeit</b> in Min.				
Arithm. Mittel	20	15	15	14
Max.-Wert	30	24	23	19
Q <sub>3</sub> = 75. PR	22	17	19	18
Med = 50. PR	20	15	15	13
(Med n. Seyfried)*	(23)	(31)	(35)	(-)
Q <sub>1</sub> = 25. PR	17	13	11	11
Min.-Wert	14	9	10	9
<b>II. Leistungsgüte</b> (Arbeitssorgfalt) Anzahl der Fehler (abs.)				
Arithm. Mittel	14	19	18	20
Max.-Wert	27	39	32	36
Q <sub>3</sub> = 75. PR	21	26	23	31
Med = 50. PR	15	18	19	19
(Med n. Seyfried)	(10)	( 6)	( 8)	(-)
Q <sub>1</sub> = 25. PR	7	12	13	10
Min.-Wert	3	6	6	5

Tabelle VI b Altersnormen zum Konzentrations-Verlaufs-Test für resthörige Schüler  
(KVT<sub>Resthör.</sub>)

KVT <sub>Resthör.</sub>	Altersgruppen			
	10;0-11;11 (N = 13)	12;0-13;11 (N = 14)	14;0-15;11 (N = 9)	16;0-17;11 (N = 7)
<b>I. Bearbeitungszeit</b> in Min.				
Arithm. Mittel	26	20	14	15
Max.-Wert	40	34	30	19
Q <sub>3</sub> = 75. PR	36	32	17	18
Med = 50. PR	24	18	15	15
(Med n. Seyfried)*	(23)	(31)	(35)	(-)
Q <sub>1</sub> = 25. PR	20	13	10	13
Min.-Wert	15	11	7	11
<b>II. Leistungsgüte</b> (Arbeitssorgfalt) Anzahl der Fehler (abs.)				
Arithm. Mittel	19	18	16	17
Max.-Wert	31	38	29	39
Q <sub>3</sub> = 75. PR	26	27	21	30
Med = 50. PR	21	19	16	18
(Med n. Seyfried)	(10)	( 6)	( 8)	(-)
Q <sub>1</sub> = 25. PR	10	10	10	3
Min.-Wert	4	1	0	1

\* Zu Vergleichszwecken sind hier die Zentralwerte (Mediane) von 35 österreichischen tauben und schwerhörigen Schülern gesondert mitangeführt (nach Seyfried; zit. bei Martin 1961, S. 31).

*Anm. zu Tab. VIa und VIb:* Bei der diagnostischen Interpretation vorstehender Normdaten ist zu beachten, daß sowohl hinsichtlich der Bearbeitungszeit als auch hinsichtlich der Fehlerzahl jeweils die Min.-Werte die beste und die Max.-Werte die schlechteste Leistung im KVT repräsentieren; analog markiert hier der Q<sub>1</sub>-Wert die obere Grenze, der Q<sub>3</sub>-Wert die untere Grenze des mittleren Leistungsbereiches, in dem 50% der Fälle liegen.

Tabelle VI c Alters-Prozentränge (PR) zum Konzentrations-Verlaufs-Test für 651 österreichische Schulkinder einschließlich 314 Internats-Sonderschüler (von H. Seyfried; zit. bei E. Martin 1961, S. 104f.) – Zeitvariable

Bearbeitungszeit des KVT in Min.	Altersgruppen			
	7;6-9;5	9;6-11;5	11;6-13;5	13;6-14;11
8	100	100	100	100
9	100	100	100	100
10	100	100	100	99
11	100	99	100	98
12	100	99	99	96
13	100	99	98	94
14	100	99	97	93
15	100	96	95	84
16	100	94	90	76
17	98	92	84	70
18	98	89	78	64
19	97	84	72	56
20	94	78	68	47
21	91	73	61	39
22	90	66	54	37
23	87	60	46	30
24	85	54	41	27
25	83	47	33	21
26	79	44	28	18
27	76	41	23	16
28	71	37	20	16
29	70	33	16	12
30	65	24	14	11
31	62	20	11	9
32	58	19	10	7
33	52	16	7	4
34	49	14	7	3
35	47	11	5	3
36	43	11	4	2
37	40	11	3	2
38	36	10	2	2
39	32	8	1	2
40	29	8	1	2
41	27	7	1	2
42	24	6	1	2
43	22	5	1	1
44	21	4	1	1
45	18	2	0	1
46	16	2	0	1
47	15	2	0	1
48	13	2	0	0
49	12	1	0	0
50	11	1	0	0
51	10	1	0	0
52	8	1	0	0
53	7	1	0	0
54	7	1	0	0
55	5	1	0	0
56	4	0	0	0
57	3	0	0	0
58	3	0	0	0
59	2	0	0	0
60	2	0	0	0
61-74	1	0	0	0
75+	0	0	0	0

Tabelle VI d Altersprozenträge (PR) zum Konzentrations-Verlaufs-Test für 1234 österreichische Schulkinder einschließlich 314 Internats-Sonderschüler (von H. Seyfried; zit. bei E. Martin 1961, S. 104 f.) – Fehlervariable

Anzahl der Fehler (abs.) im KVT	Altersgruppen							
	7;6-9;5		9;6-11;5		11;6-13;5		13;6-14;11	
	Kn	Md	Kn	Md	Kn	Md	Kn	Md
0	100	99	98	96	92	90	96	94
1	95	95	95	86	83	80	91	85
2	89	90	88	77	74	68	82	73
3	85	86	80	68	64	62	73	67
4	80	76	73	56	59	54	64	56
5	73	73	67	47	50	47	55	52
6	69	67	60	42	40	44	47	45
7	66	62	53	37	33	36	41	39
8	59	56	48	35	29	31	38	35
9	52	51	39	32	24	26	31	30
10	46	46	37	24	22	21	25	25
11	42	43	34	21	19	19	21	24
12	37	35	28	18	13	16	20	20
13	33	29	25	15	11	14	19	19
14	31	26	21	12	10	11	15	16
15	26	24	19	10	9	9	13	12
16	25	21	15	7	8	7	10	9
17	24	19	15	5	7	6	8	8
18	21	16	13	5	6	5	7	6
19	20	12	11	4	4	4	6	2
20	17	11	9	4	4	4	4	1
21	14	11	6	4	3	3	4	1
22	11	10	5	4	3	3	4	1
23	9	9	4	4	2	2	4	1
24	8	6	3	4	2	2	2	0
25	8	5	3	2	2	1	1	0
26	8	3	3	2	2	1	1	0
27	8	3	2	2	2	1	1	0
28	7	3	1	1	2	1	1	0
29	5	3	1	1	2	1	1	0
30	4	3	1	1	1	1	0	0
31	4	3	1	1	1	1	0	0
32	4	2	1	0	1	1	0	0
33	4	2	1	0	1	1	0	0
34	3	2	1	0	1	1	0	0
35	2	0	1	0	1	0	0	0
36	1	0	0	0	1	0	0	0
37	1	0	0	0	0	0	0	0
38	1	0	0	0	0	0	0	0
39	1	0	0	0	0	0	0	0
40	1	0	0	0	0	0	0	0
41	1	0	0	0	0	0	0	0
42	1	0	0	0	0	0	0	0
43+	0	0	0	0	0	0	0	0

Abk.: Kn = Knaben, Md = Mädchen

Tabelle VII a    Altersnormen zum Pauli-Test für taube Schüler (nach E. Martin 1961)

Pauli-Test	Altersgruppen			
	10;0-11;11 (N = 16)	12;0-13;11 (N = 18)	14;0-15;11 (N = 20)	16;0-17;11 (N = 6)
<b>I. Leistungsmenge</b> (Anzahl der Additionen)				
Arithm. Mittel	655	1007	1435	1315
Max.-Wert	1195	2104	2050	1721
Q <sub>3</sub> = 75. PR	956	1456	1838	1582
Med = 50. PR	600	1075	1360	1300
Q <sub>1</sub> = 25. PR	391	636	1111	1010
Min.-Wert	179	507	707	801
<b>II. Leistungsgüte</b> (Arbeitsqualität)				
1. Fehler in %				
Arithm. Mittel	6,9	5,2	2,4	2,2
Max.-Wert	57,7	31,0	16,8	8,2
Q <sub>3</sub> = 75. PR	5,3	6,7	2,7	4,8
Med = 50. PR	2,5	2,7	1,0	1,3
Q <sub>1</sub> = 25. PR	0,8	0,8	0,4	0,5
Min.-Wert	0,0	0,2	0,0	0,1
2. Verbesserungen in %				
Arithm. Mittel	2,1	2,3	2,0	1,5
Max.-Wert	11,7	5,6	5,0	3,7
Q <sub>3</sub> = 75. PR	4,5	3,1	2,1	2,7
Med = 50. PR	1,8	2,2	1,6	1,3
Q <sub>1</sub> = 25. PR	0,9	1,3	1,0	0,4
Min.-Wert	0,4	0,7	0,2	0,3

*Anm.:* Bei der diagnostischen Interpretation vorstehender Normdaten ist zu beachten, daß bezüglich des Fehler-% die Min.-Werte die beste und die Max.-Werte die schlechteste Leistung repräsentieren; analog markiert hier (im Fehler-%) der Q<sub>1</sub>-Wert die obere Grenze, der Q<sub>3</sub>-Wert die untere Grenze des mittleren Bereiches, in dem 50% aller Fälle liegen.

Tabelle VII b    Altersnormen zum Pauli-Test für taube Jugendliche (von E. Gruhnwald und E. Ulich; zit. nach W. Arnold, Der Pauli-Test. München 1961, S. 147)

Pauli-Test	10-11jährige (N = 34)	12-13jährige (N = 39)	14-15jährige (N = 38)
<b>I. Leistungsmenge</b> (Anzahl der Additionen)			
Max.-Wert	—	—	—
Q <sub>3</sub> = 75. PR	722	1476	1730
Med = 50. PR	586	994	1487
Q <sub>1</sub> = 25. PR	458	647	1048
Min.-Wert	—	—	—
<b>II. Leistungsgüte</b> (Arbeitsqualität)			
1. Fehler in %			
Max.-Wert	—	—	—
Q <sub>3</sub> = 75. PR	4,7	2,8	2,0
Med = 50. PR	1,8	0,8	0,8
Q <sub>1</sub> = 25. PR	0,9	0,4	0,4
Min.-Wert	—	—	—
2. Verbesserungen in %			
Max.-Wert	—	—	—
Q <sub>3</sub> = 75. PR	2,3	2,2	2,4
Med = 50. PR	1,6	1,4	1,6
Q <sub>1</sub> = 25. PR	0,7	0,9	1,8
Min.-Wert	—	—	—

*Anm.:* Bei der diagnostischen Interpretation vorstehender Normdaten ist zu beachten, daß bezüglich des Fehler-% die Min.-Werte die beste und die Max.-Werte die schlechteste Leistung repräsentieren; analog markiert hier (im Fehler-%) der Q<sub>1</sub>-Wert die obere Grenze, der Q<sub>3</sub>-Wert die untere Grenze des mittleren Bereiches, in dem 50% aller Fälle liegen.

Tabelle VII c Altersnormen zum Pauli-Test für resthörige Schüler (nach E. Martin 1961)

Pauli-Test	Altersgruppen			
	10;0-11;11 (N = 10)	12;0-13;11 (N = 14)	14;0-15;11 (N = 9)	16;0-17;11 (N = 5)
<b>I. Leistungsmenge</b>				
(Anzahl der Additionen)				
Arithm. Mittel	553	907	1360	1507
Max.-Wert	1150	1918	2027	1878
Q <sub>3</sub> = 75. PR	666	1200	1868	1857
Med = 50. PR	520	775	1225	1525
Q <sub>1</sub> = 25. PR	382	495	915	1085
Min.-Wert	102	200	847	954
<b>II. Leistungsgüte</b>				
(Arbeitsqualität)				
1. Fehler in %				
Arithm. Mittel	10,4	4,5	1,9	1,8
Max.-Wert	37,8	32,5	8,7	5,9
Q <sub>3</sub> = 75. PR	16,5	4,1	1,7	4,6
Med = 50. PR	6,0	1,6	1,2	0,8
Q <sub>1</sub> = 25. PR	1,2	0,9	0,7	0,5
Min.-Wert	0,4	0,3	0,4	0,4
2. Verbesserungen in %				
Arithm. Mittel	2,4	2,2	1,5	2,5
Max.-Wert	4,9	5,2	2,7	6,4
Q <sub>3</sub> = 75. PR	4,3	3,6	1,7	5,7
Med = 50. PR	2,0	2,3	1,6	1,3
Q <sub>1</sub> = 25. PR	0,5	1,2	1,1	0,5
Min.-Wert	0,3	0,2	0,8	0,4

**Anm.** Bei der diagnostischen Interpretation vorstehender Normdaten ist zu beachten, daß bezüglich des Fehler-% die Min.-Werte die beste und die Max.-Werte die schlechteste Leistung repräsentieren; analog markiert hier (im Fehler-%) der Q<sub>1</sub>-Wert die obere Grenze, der Q<sub>3</sub>-Wert die untere Grenze des mittleren Bereiches, in dem 50% aller Fälle liegen.

Tabelle VIII Transformation einschlägiger Testnormen und Zensuren

SN <sub>6</sub> <sup>1</sup>	SN <sub>5</sub> <sup>2</sup>	PR <sup>3</sup>	T <sup>4</sup>	C <sup>5</sup>	WP <sup>6</sup>	IQ <sup>7</sup>	SN <sub>6</sub>	SN <sub>5</sub>	PR	T	C	WP	IQ
		0,1	20	-1	1	55	3,5	3	50	50	5	10	100
		0,2	21	-0,8	1,5	.	.	.	54	51	5,2	.	.
		0,3	22	-0,6	2	58	.	.	58	52	5,4	10,5	103
		0,4	23	-0,4	.	.	.	.	62	53	5,6	.	.
		0,5	24	-0,2	2,5	61	.	.	66	54	5,8	11	106
5		0,6	25	0	.	.	3	2,5	69	55	6	11,5	.
.		0,8	26	0,2	3	64	.	.	73	56	6,2	12	109
.		1	27	0,4	.	.	.	.	76	57	6,4	.	.
.		1,5	28	0,6	3,5	67	.	.	79	58	6,6	12,5	112
.		2	29	0,8	.	.	.	.	82	59	6,8	.	.
5,5	5	2,5	30	1	4	70	2,5	2	84	60	7	13	115
.	.	3	31	1,2	4,5	.	.	.	86	61	7,2	.	.
.	.	3,5	32	1,4	5	73	.	.	88	62	7,4	13,5	118
.	.	4	33	1,6	.	.	.	.	90	63	7,6	.	.
.	.	5	34	1,8	5,5	76	.	.	92	64	7,8	14	121
5	4,5	7	35	2	.	.	2	1,5	93	65	8	.	.
.	.	8	36	2,2	6	79	.	.	95	66	8,2	14,5	124
.	.	10	37	2,4	.	.	.	.	96	67	8,4	.	.
.	.	12	38	2,6	6,5	82	.	.	96,5	68	8,6	15	127
.	.	13	39	2,8	.	.	.	.	97	69	8,8	15,5	.
6,5	4	16	40	3	7	85	1,5	1	97,5	70	9	16	130
.	.	18	41	3,2	.	.	.	.	98	71	9,2	.	.
.	.	21	42	3,4	7,5	88	.	.	98,5	72	9,4	16,5	133
.	.	24	43	3,6	.	.	.	.	99	73	9,6	.	.
.	.	27	44	3,8	8	91	.	.	99,2	74	9,8	17	136
4	3,5	31	45	4	8,5	.	1	.	99,4	75	10	.	.
.	.	34	46	4,2	9	94	.	.	99,5	76	10,2	17,5	139
.	.	38	47	4,4	.	.	.	.	99,6	77	10,4	.	.
.	.	42	48	4,6	9,5	97	.	.	99,7	78	10,6	18	142
.	.	46	49	4,8	.	.	.	.	99,8	79	10,8	18,5	.
3,5	3	50	50	5	10	100	.	.	99,9	80	11	19	145

Abk.: <sup>1</sup> SN<sub>6</sub> = Schulnoten (6stufige Skala); <sup>2</sup> SN<sub>5</sub> = Schulnoten (5stufige Skala);<sup>3</sup> PR = Prozentränge (cum f %); <sup>4</sup> T = T-Standardskala (M = 50, s = 10);<sup>5</sup> C = Centil-Skala (M = 5, s = 2); <sup>6</sup> WP = Wertpunkt-Skala (M = 10, s = 3);<sup>7</sup> IQ = Abweichungs-IQ (Intelligenzquotienten) oder Wechsler-IQ bzw. HAWIK-IQ (M = 100, s = 15).



T-TABELLE: RANGREIHEN-UMWANDLUNG IN "T"-STANDARDWERT von F. Tremel

Zahlen in der linken Kolonne bezeichnen die Gruppengröße (Gg)																																																																																									
Ordnungszahlen in der untersten Zeile sind Rangplätze (Rp)																																																																																									
So erhält ein "Spitzenkandidat" mit dem 2. Rangplatz in einer Bewerbergruppe von N = 36 einen linearen Skore von T = 66)																																																																																									
10	67	60	57	54	51	49	46	43	39	33	21	69	64	61	59	58	56	55	53	52	51	50	49	47	46	45	44	43	42	40	38	35	31	26	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29																						
11	67	61	57	54	52	50	48	45	42	38	32	22	69	64	62	60	58	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	27	69	65	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29																		
12	67	61	58	55	53	51	49	48	45	42	38	32	23	69	64	62	60	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	28	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29																
13	67	62	59	56	54	52	50	48	46	44	41	37	32	24	69	64	61	59	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	41	39	37	35	30	29	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29																
14	67	62	59	56	54	52	51	49	47	45	43	40	37	32	25	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	30	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29													
15	67	62	59	57	55	53	51	50	48	46	43	40	37	32	26	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	31	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29													
16	68	63	60	57	55	54	52	51	49	47	46	44	42	40	37	32	27	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	32	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29											
17	68	63	60	58	56	54	53	51	50	48	47	45	44	42	40	37	32	28	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	33	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29										
18	68	63	60	58	56	55	53	52	50	49	48	46	45	43	41	39	36	31	29	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	34	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29									
19	68	64	61	59	57	55	54	52	51	50	48	47	46	44	43	41	39	36	31	30	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	35	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29								
20	68	64	61	59	57	56	54	53	52	50	49	48	47	45	44	42	40	38	36	31	31	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	36	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29							
21	69	64	61	59	58	56	55	53	52	51	50	49	47	46	45	43	42	40	38	35	31	32	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	37	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29						
22	69	64	62	60	58	56	55	54	53	51	50	49	48	47	46	44	43	42	40	38	35	31	33	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	38	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29					
23	69	64	62	60	58	57	55	54	53	52	51	50	49	48	46	45	44	43	41	40	38	35	31	34	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	39	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29				
24	69	65	62	60	58	57	56	55	53	52	51	50	49	48	47	46	45	44	43	41	39	37	35	30	30	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	40	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29			
25	69	65	62	60	59	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	39	37	35	30	31	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	41	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29	
26	69	65	62	61	59	58	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	39	37	34	30	32	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	42	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29	
27	69	65	63	61	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	39	37	34	30	33	69	65	62	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	35	31	43	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29
28	70	66	63	61	60	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	40	38	37	34	30	34	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29																											
29	70	66	63	61	60	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	38	36	34	30	35	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	29																										
30	70	66	63	62	60	59	57	56	55	54	53	52	51	50	49	48	47	46	45	44	43	42	41	40	38	36	34	30	36	70	66	63	61	60	59	58	57	56	55	54	53	52	51	50	49	48	47	46																																									

## Zusammenfassung

Das Buch besteht aus einem theoretischen und einem praktisch-empirischen Teil. Zunächst werden Ergebnisse der neueren Begabungsforschung erörtert. Dabei kommen sowohl phänomenologisch orientierte Theoretiker (z.B. Wenzl, Gottschaldt, Mierke) als auch die für die Intelligenzdiagnostik bedeutsamen Faktorenmodelle von Spearman et al., Meili, Thurstone, Guilford und Jäger zur Sprache. Eingehend werden milieutheoretische und psycholinguistische Forschungsansätze im Kontext begabungsdiagnostischer Fragestellungen behandelt. Schließlich werden familiäre und schulpädagogische Konsequenzen für eine optimale intellektuelle Förderung im Kindes- und Jugendalter diskutiert, wobei dem Zusammenhang von Sprache und Begabung besondere Beachtung geschenkt wird.

Das zweite Kapitel informiert über grundlegende Voraussetzungen intelligenzdiagnostischer Untersuchungen und dient der Einführung in die Testtheorie unter besonderer Berücksichtigung der Testgütekriterien bzw. klassischen Meßfehlertheorie. Diese Ausführungen wenden sich, genauso wie das nachfolgende Kapitel, in erster Linie an Pädagogen resp. diagnostisch tätige Lehrer und Sozialwissenschaftler. Deshalb werden zunächst Ziele und Aufgaben der Testanwendung im schulischen Bereich formuliert. Mittelpunkt der Überlegungen zum Thema Schülerbeurteilung und Bildungsberatung bilden dann Erörterungen zum Design der Datenerfassung (Test, Verhaltensbeobachtung, Fragebogentechnik, Anamnese und Exploration) und Gutachtenerstellung. Eine Reihe konkreter Fallbesprechungen soll dem Leser den oft als schwierigsten Schritt empfundenen Vorgang der Datenintegration versus Befundinterpretation veranschaulichen und somit zu einem besseren Verständnis diagnostischer Aussagen oder auch einer sichereren Urteilspraxis beitragen, zumal brauchbare Hinweise hierzu in den meisten Diagnostiklehrbüchern fehlen.

Im zweiten (empirischen) Teil des Buches finden sich zahlreiche Forschungsergebnisse über einschlägige Probleme der Begabungsdiagnostik innerhalb der Regel- und Sonderschule. Neben einem ausführlicher gehaltenen Methodenreferat (über Intelligenztests, Verfahren zur Erfassung der Leistungsmotivation, Konzentration, Interessen etc.) stehen hier zentrale Probleme pädagogischer Intelligenzbeurteilung, etwa im Rahmen von Schulleistungsanalysen, Probleme und Methoden der Schuleignungsermittlung am Ende der Grundschule oder auf den weiterführenden Schulen (Hauptschule, Realschule, Gymnasium), aber auch Grundsatzprobleme wie die Interpretation von Testprofilen zur Diskussion. Daneben werden intelligenzdiagnostische Probleme und bisher unveröffentlichte Ergebnisse zur Lernbehinderten-, Hör- und Sehgeschädigtendiagnostik besprochen. Ein 25seitiger Tabellenanhang mit zahlreichen Testnormen für hör/sprachbehinderte Kinder und Jugendliche sowie ein kurzer Ausblick auf vordringliche Forschungsvorhaben beschließen das Werk, das nicht nur Informationen vermitteln, sondern gleichzeitig einen Beitrag zur Umsetzung einschlägiger Forschungsergebnisse in die pädagogisch-diagnostische Praxis leisten möchte.

## Summary

The book comprises a theoretical and a practical-empirical section. Firstly results of recent research on the assessment of intellectual abilities are reviewed and reference is made to phenomenological oriented theorists like Wenzl, Gottschaldt, Mierke on the one hand and to factoranalytical oriented theorists like Spearman et al., Meili, Thurstone, Guilford and Jäger on the other hand, who have essentially contributed to the assessment of intellectual abilities. Within the framework of current issues on the topics pertaining to assessment of talent, reference is made to research issues implicating socialization and psycholinguistic aspects. Finally discussion is centered on behavioral consequences for parents and teachers in creating situational attributes that will have most beneficial effects on the development of intellectual abilities in the early stages of life and in this context the interdependent relationship of language and intellectual development is stressed.

The second chapter informs on basic prerequisites in assessment of intelligence and is intended to introduce the reader into theoretical backgrounds of testing. Special consideration is given to the topics of validity, reliability and psychometric standards. The essentials on these topics, together with the following chapter, are addressed mainly to those working in education and social science who are concerned with the application of psychological tests. This is why firstly the range of test application in the area of education is discussed. In the course of the following thoughts on evaluation of aspects of pupils personality and counseling on educational possibilities for the individual discussion is centered on design of data collection like choice of tests, observational procedures of behavior, aspects of personality inventories, techniques of assessment of biographical data and the technique of reporting the results. Quite a few empirical case reports are intended to help the reader to overcome the often complained of difficult barrier between the integration of data vs. the interpretation of the findings. This process is frequently experienced as difficult by many practitioners and it is hoped, the illustration of case material will contribute to a better understanding of diagnostic reports and an adequate evaluation of diagnostic statements and help the reader in acquiring higher proficiency in evaluating personality and ability characteristics, problems that are frequently neglected in basic books on diagnostics.

The second (empirical) part of the book highlights quite a few research findings on current problems in the area of evaluation of abilities in primary and secondary schools or in schools for special education, e.g. schools for mentally retarded or children with learning inhibitions, pupils with hearing and speech handicaps etc. Together with an intensive discussion on methodological issues (like measuring intelligence, achievement motivation, vigilance, attention and interests) interest is focussed on the essentials of evaluation of intellectual abilities as related to educational aspects, e.g. on the background of analysis of achievements in school, problems and methods of measuring qualification for further education on leaving primary school or on subsequent levels (highschool etc.); basic problems like the interpretation of test profiles are likewise advanced. In addition there is a review of problems on the assessment of intellectual functioning and – up till now – unpublished findings on diagnostic testing of pupils with decrements in learning capacity and reduced auditory and visual functioning levels. Included at the end of the book is a 25-page appendix of tables with numerous test norms for children and youngsters with hearing and speech handicaps. A short perspective on research issues which should be given priority concludes the book, which is intended not only to give broad information, but also to contribute to a more effective integration of current research findings into practical diagnostic work in the area of education.

## Résumé

Le livre se compose d'une partie théorique et d'une partie pratique empirique. D'abord on discute les résultats de la récente recherche concernant l'intelligence. On y parle aussi bien de théoriciens orientés par la phénoménologie (par exemple Wenzl, Gottschaldt, Mierke) que des modèles de facteurs de Spearman et al., Meili, Thurstone, Guilford et Jäger qui sont importants pour le diagnostic de l'intelligence. Les dispositions de recherche fondées sur la théorie de milieu et la psycho-linguistique s'y trouvent traitées en détail dans le contexte de problèmes concernant le diagnostic de l'intelligence. Enfin il s'ajoute une discussion au sujet des conséquences familiales et scolaires ayant pour but un maximum d'encouragement intellectuel dans l'enfance et la jeunesse, discussion au cours de laquelle on accorde une attention particulière au rapport qui existe entre la langue et l'intelligence.

Le deuxième chapitre informe des données fondamentales que demandent les études examinant le diagnostic de l'intelligence et il sert en même temps à l'introduction à la théorie de test tout en tenant compte particulièrement des critères des tests ou de la théorie des fautes. Ces explications de même que le chapitre suivant, s'adressent d'abord aux pédagogues. C'est pourquoi on formule d'abord les buts et les fonctions de l'application des tests dans le domaine scolaire. Au centre des réflexions faites au sujet des jugements d'élèves et du conseil d'éducation se trouvent des remarques portant sur la forme de la perception de dates (tests, observation de conduite, technique du questionnaire, anamnèse, exploration) et sur la rédaction de jugements. Une série d'analyses de cas concrets doit illustrer au lecteur le passage de l'intégration à l'interprétation des dates, procédé, compté souvent parmi les plus difficiles, et elle doit ainsi contribuer à une meilleure compréhension ou à une pratique plus sûre des jugements diagnostiques puisque la plupart des manuels de diagnostic sont insuffisants à cet égard.

Dans la deuxième partie du livre, partie empirique, se trouvent de nombreux résultats de recherche portant sur les problèmes principaux du diagnostic de l'intelligence à l'intérieur de l'école ou de l'école spéciale. À côté d'un exposé plus détaillé traitant les méthodes (les tests de l'intelligence, les procédés servant à explorer la motivation du travail, la concentration, les intérêts etc.) on a mis en discussion ici les problèmes centraux du jugement pédagogique de l'intelligence, par exemple dans le cadre des analyses du travail scolaire, des problèmes et des méthodes concernant l'examen de l'aptitude scolaire à la fin de l'école élémentaire ou dans une autre école (Hauptschule, Realschule, Gymnasium) mais aussi des problèmes de base comme l'interprétation de tests. En outre on discute des problèmes de diagnostic de l'intelligence et des résultats, pas encore publiés jusque-là, qui concernant le diagnostic des débiles mentaux et des handicapés d'ouïe ou de langue ou de vue. Le livre se termine sur un appendice de tableaux de 25 pages offrant de nombreuses normes de tests faites pour les enfants et les jeunes handicapés d'ouïe et de langue et sur une petite perspective exposant les projets de recherche de première urgence. Le livre ne veut pas seulement communiquer des informations, mais il veut en même temps contribuer à mettre en pratique, dans le domaine du diagnostic pédagogique, les résultats de recherche de grande importance.

Книга содержит теоретическую и конкретную часть .

Прежде всего излагаются результаты новейших исследований способности. В связи с этим обсуждаются феноменологически направленные исследователи - теоретики как Wenzl, Gottschaldt, Mierke, как и важные модели факторов Spearman, Meili, Thurston Guilford и Jaeger.

Наряду с вопросами о определении способностей обсуждаются исследовательские начала в контексте с теорией окружения и психолингвистики.

Наконец обсуждаются семейные и школьно-педагогические выводы для оптимального умного поощрения детей и подростков, причем особенное внимание уделяется содействию языка и умственных способностей.

Вторая глава содержит принципиальные предпосылки для исследования умственной способности и служит как введение теорию опыта причем особенное внимание уделяется точности опыта.

Эти выводы сделанны прежде всего для педагогов и для исследователей общественных наук. Потому определяется цели и задачи применения опыта в школах. В связи с аттестованием ученика и уделением советов для дальнейшего образования служат с одной стороны : опыт, наблюдение поведения, анкеты, биографические данные и с другой стороны : экспертиза.

Конкретны примеры используются для объяснения выводов экспериментов чтобы таким образом облегчить педагогам аттестирование учеников.

Вторая часть содержит целый ряд примеров исследований современных проблем в области определения умной способностей учеников общих и чрезвычайных школ.

Центральные проблемы педагогического определения умственных способностей находятся в центре внимания наряду с обзором методов работы так на пример : опыты умственных способностей, определение степени стремления к успехам.

Эти опыты применяются у выпускников обще образовательных и средних школ.

В прочем обсуждаются до сих пор еще не опубликованные итоги для определения умственных способностей у учеников способностями зрения и слушания.

Книга содержит кроме того 25 страниц с нормами для проведения опытов.

# Literaturverzeichnis

- Abels, D., 1961. Der Konzentrations-Verlaufs-Test (K-V-T). 2. Aufl., Göttingen
- Aebli, H., 1969. Die geistige Entwicklung als Funktion von Anlage, Reifung, Umwelt- und Erziehungsbedingungen. In: Begabung und Lernen (Hrsg. H. Roth). 3. Aufl., Stuttgart, 151-191
- Affolter, F., 1953. Über symbolische, anschauliche und logisch-konkrete Denkabläufe bei taubstummen und hörenden Kleinkindern und Schulanfängern nach Piaget. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Allinger, U. u. Heller, K., 1972. Automatische Klassifikation von psychologischen Untersuchungsbefunden. In: Bildung in neuer Sicht (= Schriftenreihe A Nr. 28 des Kultusministeriums Baden-Württemberg zur Bildungsforschung, Bildungsplanung, Bildungspolitik). Villingen
- Amelang, M. u. Zimmermann, K.W., 1966. Versuch einer intelligenzdiagnostischen Differenzierung zwischen Volksschülern und lernbehinderten Sonderschülern. Marburg (unveröffentl. Manuskript.)
- Amelang, M. u. Zimmermann, K.W., 1968. Die Faktorenstruktur des HAWIK bei schwachbegabten Kindern. Heilpäd. Forschg., 1, 381-389
- Amthauer, R., 1955. Der Intelligenz-Struktur-Test (I-S-T). 2. Aufl., Göttingen
- Anastasi, A., 1958. Differential Psychology. New York
- Anastasi, A., 1968. Psychological Testing. 3. Aufl., New York
- Anderson, H.H. (Hrsg.), 1959. Creativity and its Cultivation. New York, Evanston
- Andre, K. u. Gebauer, E., 1955. Vergleichende Untersuchungen über das Verhalten und die Leistungen acht- bis zehnjähriger taubstummer und hörender Schüler beim Lösen von anschaulich gegebenen Problemen. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Anger, H. u. Bargmann, R., 1954. Entwicklung und Eichung der Frankfurter Wortschatztests. Göttingen
- Anger, H.; Bargmann, R.; Hylla, E., 1965a. Wortschatztest WST 5-6. Begabungstest für 5. und 6. Klassen. Neubearbeitet von H. Horn. Weinheim
- Anger, H.; Bargmann, R.; Hylla, E., 1965b. Wortschatztest WST 7-8. Begabungstest für 7. und 8. Klassen. Neubearbeitet von H. Horn. Weinheim
- Arnold, W., 1961. Der Pauli-Test. Anweisung zur sachgemäßen Durchführung, Auswertung und Anwendung des Kraepelinschen Arbeitsversuches. 3. Aufl., München
- Atkinson, J.W. (Hrsg.), 1958. Motives in Fantasy, Action and Society. New York
- Aurin, K., 1966. Ermittlung und Erschließung von Begabungen im ländlichen Raum – Untersuchung zur Bildungsberatung in Baden-Württemberg. Villingen
- Aurin, K. et al., 1968. Gleiche Chancen im Bildungsgang. (= Schriftenreihe A Nr. 9 des Kultusministeriums Baden-Württemberg zur Bildungsforschung, Bildungsplanung, Bildungspolitik). Villingen
- Baitinger, U. u. Bernd, R., 1967. HAWIK-Untersuchungen bei blinden und sehbehinderten Schülern. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Baitinger, U. u. Bernd, R., 1970. HAWIK-Untersuchungen bei Blinden und Sehbehinderten. Z. f. Blindenbildungswesen (Der Blindenfreund), 90, 134-142.
- Bangen, R., 1965. Untersuchungen der Intelligenz gehörloser Kinder und Jugendlicher mit dem Progressive-Matrices-Test von J.C. Raven. Examensarbeit am Sonderpäd. Inst. Hamburg (unveröffentl.)
- Barron, F., 1965. The Psychology of Creativity. In: New Directions in Psychology, II (Hrsg. T.M. Newcomb). New York, 1-134
- Bartenwerfer, H., 1964. Allgemeine Leistungstests. In: Hb d. Psychol., Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß). Göttingen, 385-410
- Batt, H., 1967. Vergleichende Intelligenzuntersuchungen an hörgeschädigten Kindern: Progressive Matrices – Benton-Test. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Baumeister, A.A. u. Bartlett, C.J., 1962 a. A comparison of the factor structure of normals and retardates on the WISC. Amer. J. ment. Defic., 66, 641-646

- Baumeister, A.A. u. Bartlett, C.J., 1962 b. Further factorial investigations of WISC performance of mental defectives. *Amer. J. ment. Defic.*, 67, 257-261
- Bäumler, G., 1964. Zur Faktorenstruktur der Paulitestleistung unter besonderer Berücksichtigung des sogenannten numerischen Faktors. *Diagnostica*, 10, 107-119
- Bäumler, G.; Seitz, W.; Knop, G.; 1968. Die faktorielle Struktur der "Nichtverbalen Intelligenztestreihe" von Snijders und Snijders-Oomen bei Taubstummen. *Diagnostica*, 14, 87-93
- Beck, U., 1969. Auslesekriterien für den Englischunterricht an Hauptschulen. Examensarbeit zur II. Dienstprüf. f. d. Volksschullehramt. PH Ludwigsburg (unveröffentl.)
- Belser, H., 1967. Testentwicklung, 2. Aufl. Weinheim, Berlin
- Belser, H.; Anger, H.; Bargmann, R., 1965 a. Frankfurter Analogietest FAT 4-6. Begabungstest für 4.-6. Klassen. Weinheim
- Belser, U.; Anger, H.; Bargmann, R., 1965 b. Frankfurter Analogietest FAT 7-8. Begabungstest für 7.-8. Klassen. Weinheim
- Benton, A.L., 1945. A visual retention test for clinical use. *Arch. Neurol. Psychiat.* 54, 212-216
- Benton, A.L., 1961. Der Benton-Test. Vorlagenhefte 1 und 2. Bern, Stuttgart
- Benton, A.L., 1968. Der Benton-Test. Handbuch. Deutsche Bearbeitung von O. Spreen 3. Aufl., Bern, Stuttgart
- Bereiter, C., 1966. A beginning language program for disadvantaged children. Paper presented at the Am. Educ. Res. Association. Chicago. (Zit. nach U. Oevermann 1969)
- Bereiter, C., o.J. The relative importance of verbal and nonverbal factors in cultural deprivation: Evidence from children with sensory handicaps. Mimeogr. Paper, Urbana, Ill., Univ. of Illinois. (Zit. nach U. Oevermann 1969)
- Bernstein, B., 1958. Some sociological determinants of perception: An enquiry into sub-cultural differences. *British J. Soc.*, 9, 159-174
- Bernstein, B., 1959. A public language: Some sociological implications of linguistic form. *British J. Soc.*, 10, 311-326
- Bernstein, B., 1962. Social class, linguistic codes and grammatical elements. *Language and Speech*, 5, 221-240
- Bernstein, B., 1964. Elaborated and restricted codes: Their origins and some consequences. In: *The Ethnography of Communication* (Hrsg. J.J. Gumperz und D. Hymes). *Am. Anthropologist*, Special Publication, 66, Nr. 6, Part. 2, 55-69
- Bernstein, B., 1967. Sozio-kulturelle Determinanten des Lernens. In: *Pädagogische Psychologie* (Hrsg. F. Weinert). Köln, Berlin, 346-371
- Binet, A. u. Simon, T., 1908. Le développement de l'intelligence chez les enfants. *Année psychol.*, 14, 1-94
- Bleidick, U., 1970. Das sonderpädagogische Gutachten. Praktische Anleitung zur Beobachtung und Beurteilung von Sonderschulkindern. 3. Aufl., Berlin
- Bloom, B.S., 1966. *Stability and Change of Human Characteristics*. 3. Aufl., New York, London, Sidney
- Bloom, B.S., 1967. Über das Testen kognitiver Fähigkeiten und Leistungen. In: *Pädagogische Psychologie* (Hrsg. F. Weinert). Köln, Berlin, 457-469
- Bloom, B.S. et al., 1965. *Compensatory Education for Cultural Deprivation*. New York
- Bondy, C. (Hrsg.), 1956. *Die Messung der Intelligenz Erwachsener*. Textband zum HAWIE. 3. Aufl., 1964, Bern, Stuttgart
- Bondy, C. (Hrsg.), 1956 a. *Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE)*. 3. Aufl. 1964, Bern, Stuttgart
- Bondy, C. (Hrsg.), 1956 b. *Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK)*. 3. Aufl. 1966, Bern, Stuttgart
- Bourdon, B., 1895. Observations comparatives sur la reconnaissance la discrimination et l'association. *Rev. philos.*, 40, 153-185.
- Bourdon, B., 1902. Recherches sur l'habitude. *Année psychol.*, 8, 327-340
- Bracken, H. v., 1965. *Entwicklungsgestörte Jugendliche*. München



- Brennecke, O., 1962. Untersuchungen der allgemeinen Intelligenz von 70 Kindern der Volks- und Mittelschule für Schwerhörige in Hamburg mit dem Snijders-Oomen-Test und dem Handlungsteil des HAWIK und Vergleich der beiden Testverfahren sowie den bei Schwerhörigen erzielten Ergebnissen. Examensarbeit am Sonderpäd. Inst. Hamburg (unveröffentl.)
- Bronfenbrenner, U., 1958. Socialization and social class through time and space. In: Readings in Social Psychology (Hrsg. E.E. Maccoby et al.). New York
- Brown, W., 1910. Some experimental results in the correlation of mental abilities. *British J. Psychol.*, 3, 296-322
- Bruner, J.S. et al., 1966. *Studies in cognitive growth*. New York
- Brunner, P.; Lichtenberger, W.; Schwinger, L., 1966. Intelligenzdiagnostische Untersuchungen mit dem Progressiven Matrizentest zur Ermittlung von Testnormen für gehörgeschädigte Kinder. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Buck, J.N., 1948. The H-T-P-Technique. A Qualitative and Quantitative Scoring Manual. *J. Clin. Psychol.*, 4, 319-396. (= Monogr. Suppl. Nr. 5 Oct. 1948. Brandon, Vermont)
- Büscher, P., 1964. Intelligenz, Schulleistung und soziales Verhalten bei taubstummen und hörenden Schulkindern. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Burt, C., 1949. The structure of the mind: a review of the results of factor analysis. *British J. Educ. Psychol.*, 19, 100-111, 176-199
- Cattell, R.B., 1944. Psychological measurement: normative, ipsative, interactive. *Psychol. Rev.*, 51, 292-303
- Cattell, R.B., 1962. Culture Free Intelligence Test, Scale 3. Form A and B. 2. Aufl., Champaign, Ill.
- Cattell, R.B. u. Cattell, A.K.S., 1957. Test of "g": Culture Fair, Scale 2. Form A and B. 2. Aufl., Champaign, Ill.
- Cattell, R.B. u. Weiß, R., 1970. Grundintelligenztest. Deutsche Bearbeitung der Culture Fair oder Culture Free Tests. Braunschweig
- Chauncey, H. u. Dobbin, J.E., 1970. Der Test im modernen Bildungswesen. 2. Aufl., Stuttgart
- Chomsky, N., 1961. Some methodological remarks on generative grammar. *Word*, 17, 219-239
- Cohen, R., 1964. Zeichentests zur Prüfung der Intelligenz. In: Hb. d. Psychol. Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß) Göttingen, 260-279
- Cooley, W.W. u. Lohnes, P.R., 1962. *Multivariate Procedures for the Behavioral Sciences*. New York, London, Sidney
- Correll, W. u. Ingenkamp, K., 1967. Fremdsprachen-Eignungstest für die Unterstufe FTU 4-6. Deutsche Parallelentwicklung zur Elementarform des Modern Language Aptitude Tests von J.B. Carroll und S.M. Sapon. Weinheim, Berlin
- Cronbach, L.J., 1970. *Essentials of Psychological Testing*. 3. Aufl., New York
- Cronbach, L.J. u. Gleser, G.C., 1965. *Psychological tests and personnel decisions*. 2. Aufl., Urbana
- Dahl, G., 1968. Übereinstimmungsvalidität des HAWIE und Entwicklung einer revidierten Testform. Meisenheim am Glan
- Dalferth, B., 1964. Wort- und Zahlengedächtnis bei blinden Schülern. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Daniels, J.C., 1962. *Figure Reasoning Test (FRT)* London
- Dautermann, W.L.; Shapiro, B.; Suinn, R.M., 1967. Performance tests of intelligence for the blind reviewed. *Intern. J. for the Educ. of the Blind*, 16, 8-16
- Deller, E., 1963. Über Begriff und Begriffsbildung (mit einer vergleichenden experimentellen Untersuchung über die Invariantenbildung bei hörenden und taubstummen Kindern des 10. und 13. Lebensjahres). Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Deutsch, C.P., 1965. Effects of environmental deprivation on basic psychological processes: Some hypotheses. Paper Pres. Annual Meeting of the Am. Educ. Res. Assn., Chicago, Ill. (Zit. nach U. Oevermann 1969)

- Deutsch, M., 1963. The disadvantaged child and the learning process: Some social psychological and developmental considerations. In: Education in Depressed Areas (Hrsg. H. Passow). New York, 163-179
- Deutscher Bildungsrat (Hrsg.), 1970. Strukturplan für das Bildungswesen. (= Empfehlungen der Bildungskommission). Bonn, Stuttgart
- Doman, G. u. Lückert, H.R., 1967. Wie kleine Kinder lesen lernen. 2. Aufl., Freiburg/Br.
- Donat, H., 1970. Persönlichkeitsbeurteilung – Methoden und Probleme der Charaktererfassung im pädagogischen Bereich. 2. Aufl., München
- Drever, G. u. Fröhlich, W.D., 1968. Wörterbuch zur Psychologie. Dtv. 3031. München
- Ebbinghaus, H., 1885. Über das Gedächtnis. Leipzig
- Ebbinghaus, H., 1897. Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. Z. Psychol., 13, 401-459
- Eckrich, W., 1970. Die Verwendbarkeit der Weinheimer (Frankfurter) Wortschatztests WST 5-6 und WST 7-8 in der Schwerhörigenschule. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Eigler, G. et al., 1971. Wissenschaftliche Begleituntersuchungen an Modellschulen (= Schriftenreihe A Nr. 23 des Kultusministeriums Baden-Württemberg zur Bildungsforschung, Bildungsplanung, Bildungspolitik). Villingen
- Estes, B.W., 1953. Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children: an exploratory study. J. consult. Psychol., 17, 58-62
- Estes, B.W., 1955. Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children: Addendum. J. consult. Psychol., 19, 225-226
- Eyferth, K., 1963. Gültigkeit und faktorielle Struktur des Hamburg-Wechsler-Intelligenztests für Kinder (HAWIK) bei seiner Anwendung auf schwachsinnige Erwachsene. Diagnostica, 9, 11-26
- Evans, L., 1966. A Comparative Study of the Wechsler Intelligence Scale for Children (Performance) and Raven's Progressive Matrices with Deaf Children. The Teacher of the Deaf, 69, 76-82
- Faber, E. u. Nollau, W., 1969 a. Über einen Algorithmus zur mehrdimensionalen Diskriminanzanalyse. Schriftenreihe des DRZ, Heft S-5, Darmstadt
- Faber, E. u. Nollau, W., 1969 b. Über ein Verfahren zur automatischen Klassifikation. Schriftenreihe des DRZ, Heft S-6, Darmstadt
- Feldt, L.S., 1967. A note on the use of confidence bands to evaluate the reliability of a difference between two scores. Am. Ed. Res. J., 139-145
- Fischer, G.H. (Hrsg.), 1968. Psychologische Testtheorie. Bern, Stuttgart
- Flechsig, K.K. et al., 1969. Die Steuerung und Steigerung der Lernleistung durch die Schule. In: Begabung und Lernen (Hrsg. H. Roth). 3. Aufl., Stuttgart, 449-503
- Floud, J., (Hrsg.) 1956. Social class and educational opportunity. London
- French, J.W., 1951. The description of aptitude and achievement tests in terms of rotated factors. Psychometric Monogr. 5
- Fröhlich, W.D., 1965. Forschungsstatistik. 4. Aufl., Bonn
- Frohn, W., 1926. Untersuchungen über das Denken der Taubstummen. Arch. ges. Psychol. 55, 459-523
- Frohn, W., 1927. Über Begriffsbildung und Wortbedeutung. Z. f. Kinderforsch., 33, 469-500
- Frohn, W., 1929. Wahrnehmung, Begriffsbildung und Sprache. Blätter f. Taubstummenbildg., 42, 97-104, 113-120, 151-159, 183-188
- Frohn, W., 1950. Die Rechenschwäche der Taubstummen im Lichte der Denkpsychologie. Neue Blätter f. Taubstummenbildg., 5, 65-102
- Furth, H.G., 1964. Research with deaf. Implications for language and cognition. Psych. Bull., 62. 145-164
- Furth, H.G., 1966. Thinking without language. New York
- Gaude, P. u. Teschner, W.P., 1970. Objektivierete Leistungsmessung in der Schule – Einsatz Informeller Tests im Leistungsdifferenzierten Unterricht. Frankfurt/M., Berlin, München

- Gaus, A., 1969. Die Begabung und Schulbildung. In: *Begabungsforschung und Bildungsförderung als Gegenwartsaufgabe* (Hrsg. H.R. Lückert). München, Basel, 171-183
- Gebauer, T., 1964. Vergleichende Untersuchung über den Voraussagewert von Aufnahmeprüfung und Testuntersuchung für den Erfolg auf weiterführenden Schulen. Weinheim
- Gebhardt, H., Intelligenzuntersuchungen an gehörlosen Schülern und Schülerinnen der Rheinischen Landesschule für Gehörlose in Wuppertal-Elberfeld (unveröffentl. Manuskript.)
- Getzels, J.W., 1966. Pre-school education. *Teachers College Record*, 68, 219-228
- Goetzing, C.P. u. Rousey, C.L., 1959. Educational achievement of deaf children. *Am. Ann. Deaf.*, 104, 221-231
- Gottschaldt, K., 1939. Erbsychologie der Elementarfunktionen der Begabung. In: *Hb. d. Erbbiologie d. Menschen* (Hrsg. G. Just), Bd. V, 1. Teil, Berlin, 445-537
- Gottschaldt, K., 1969. Begabung und Vererbung. Phänogenetische Befunde zum Begabungsproblem. In: *Begabung und Lernen* (Hrsg. H. Roth), 3. Aufl., Stuttgart, 129-150
- Graumann, C.F., 1955. Die Kriterien des Einfallslebens – Eine theoretische und experimentelle Analyse. Köln (Diss. 1952)
- Graumann, C.F., 1960. Eigenschaften als Problem der Persönlichkeitsforschung. In: *Hb. d. Psychol.*, Bd. 4: Persönlichkeitsforschung und Persönlichkeitstheorie (Hrsg. P. Lersch u. H. Thomae). Göttingen, 87-154
- Graumann, C.F., (Hrsg.), 1965. *Denken*. 4. Aufl. 1969, Köln, Berlin
- Graumann, C.F., 1966. Grundzüge der Verhaltensbeobachtung. In: *Fernsehen in der Lehrerbildung* (Hrsg. E. Meyer). München, 86-107
- Gray, S.W. u. Klaus, R.A., 1966. Deprivation, development, and diffusion. *Papers and Reports on the Demonstration and Research Center for Early Education*, 1, Nr. 1 (George Peabody College for Teachers, Nashville, Tenn.)
- Groffmann, K.J., 1964. Die Entwicklung der Intelligenzmessung. In: *Hb. d. Psychol.*, Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß). Göttingen, 148-199
- Guilford, J.P., 1956. The structure of intellect. *Psychol. Bull.*, 53, 267-293
- Guilford, J.P., 1959 a. Personality. New York – Deutsch: Persönlichkeit. 2/3 Aufl. 1965, Weinheim
- Guilford, J.P., 1959 b. Traits of Creativity. In: *Creativity and its Cultivation* (Hrsg. H.H. Anderson). New York, Evanston, 142-161
- Guilford, J.P., 1964. Progress in the Discovery of Intellectual Factors. In: *Widening Horizons in Creativity* (Hrsg. C.W. Taylor). New York etc., 261-297
- Guilford, J.P. u. Hoepfner, R., 1971. *The Analysis of Intelligence*. New York
- Gulliksen, H., 1950. *Theory of Mental Tests*. New York
- Hammond, K.R. u. Allan, J.M., 1953. *Writing clinical reports*. New York
- Hardesty, F.P. u. Priester, H.J., 1966. *Handbuch für den Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK)*. 3. Aufl., Bern, Stuttgart
- Hartmann, H., 1970. *Psychologische Diagnostik*. Urban-Taschenb., Bd. 135. Stuttgart usw.
- Hasemann, K., 1964. Verhaltensbeobachtung. In: *Hb. d. Psychol.*, Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß). Göttingen, 807-836
- Hayes, S.P., 1941. *Contributions to a psychology of blindness*. New York
- Hayes, S.P., 1962. Measuring the Intelligence of the Blind. In: *Zahl, P.A. (Hrsg.), Blindness*. New York, 141-173
- Heckhausen, H. 1966. Einflüsse der Erziehung auf die Motivationsgenese. In: *Psychologie der Erziehungsstile* (Hrsg. T. Herrmann). Göttingen, 131-169
- Heckhausen, H., 1969. Förderung der Lernmotivierung und der intellektuellen Tüchtigkeiten. In: *Begabung und Lernen* (Hrsg. H. Roth) 3. Aufl., Stuttgart, 193-228
- Heider, F. u. Heider, G.M., 1940. A comparison of sentence structure of deaf and hearing children. In: *Studies in Psychology of the Deaf*, 1, *Psychol. Monographs*, 52, Whole No. 232, 42-103. (Zit. nach U. Oevermann 1969)
- Heiß, R., 1960. Zum Begriff der Intelligenz. *Diagnostica*, 6, 3-11
- Heiß, R. (Hrsg.), 1964. *Psychologische Diagnostik*. (= *Hb. d. Psychologie*, Bd. 6). Göttingen
- Heiß, R., 1964. Technik, Methodik und Problematik des Gutachtens. In: *Psychol. Diagnostik*, Bd. 6 d. *Hb. d. Psychol.* (Hrsg. R. Heiß). Göttingen, 975-995

- Heller, K., 1965. Wissenswertes über psychologische Testverfahren. Neue Blätter f. Taubstummenbildg., 19, 179-190
- Heller, K., 1966. Voraussetzungen und Möglichkeiten moderner wissenschaftlicher Forschung. Neue Blätter f. Taubstummenbildg., 20, 162-174, 290-308
- Heller, K., 1967 a. Leistungsbeurteilung in der Gehörlosenschule. In: Enzyklopäd. Hb. d. Sonderpädagogik (Hrsg. G. Heese u. H. Wegener). Berlin, 2000 bis 2002
- Heller, K., 1967 b. PMT-Normen für hörgeschädigte Kinder. Neue Blätter für Taubstummenbildg., 21, 226-234
- Heller, K., 1968. Begabungsbestand in Baden-Württemberg. Forschungsbericht für das Kultusministerium Baden-Württemberg (unveröffentl.)
- Heller, K., 1969 a. Bedingungsvariablen der Schulleistung. Neuere Forschungsergebnisse und ihre Konsequenzen für die familiale und schulische Bildungsarbeit. Neue Blätter f. Taubstummenbildg., 23, 162-172
- Heller, K., 1969 b. Zum Problem der Begabungsreserven. In: Begabungsforschung und Bildungsförderung als Gegenwartsaufgabe (Hrsg. H.R. Lückert). München, Basel, 352-430
- Heller, K., 1970 a. Aktivierung der Bildungsreserven. Bern, Stuttgart
- Heller, K., 1970 b. Psychologische Untersuchungen zur Erfassung der Schuleignungsreserven. Z. f. Entwicklungspsychol. u. Päd. Psychol., 2, 223-240
- Heller, K., 1972 a. Entwicklungsstörungen der Sprache im Kindes- und Jugendalter. Heilpäd. Forschg., 4, 35-56
- Heller, K., 1972 b. Sehgeschädigtenpsychologie als Wissenschaftsdisziplin. In: Sehgeschädigte – Intern. Wiss. Arch., 1, 50-64
- Heller, K., 1972 c. Untersuchung zur Schuleignungsermittlung in Mannheim. In: Bildung in neuer Sicht (= Schriftenreihe A Nr. 28 des Kultusministeriums Baden-Württemberg zur Bildungsforschung, Bildungsplanung, Bildungspolitik). Villingen
- Heller, K., 1973. Die Bedeutung der Zeitvariablen für die Entwicklung von Begabung und Lernleistung (Schulleistung). In: Das Zeitproblem im Lernprozeß (Hrsg. W. Breunig). München
- Heller, K.; Demel, E.; Schorre, G., 1969. Modell eines Guidance-Systems für Abiturienten und Studenten. In: Forschungsergebnisse und Materialien zum Hochschulgesamtplan I Baden-Württemberg (= Schriftenreihe A Nr. 20 des Kultusministeriums Baden-Württemberg zur Bildungsforschung, Bildungsplanung, Bildungspolitik). Villingen, 109-162
- Heller, K., u. Köhn, B., 1973. Wortschatztests für Blinde – Punktschriftausgabe WST (BI 4-6) und WST (BI 7-8+), Weinheim, Berlin
- Heller, K., u. Schirmer, B., 1973. Wortschatztests für Sehbehinderte – Großdruckausgabe WST (Sb 4-6) und WST (Sb 7-9). Weinheim, Berlin
- Hengstler, M., 1972. Faktorenanalytische Untersuchungen der HAWIK-Leistung Sehgeschädigter. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Herderschêe, D., 1920. Teste für Taubstumme. Z. f. angew. Psychol., 16, 40-61
- Herrmann, T. (Hrsg.), 1966. Psychologie der Erziehungsstile. Beiträge und Diskussionen des Braunschweiger Symposions vom 28.3. bis 31.3.1966. Göttingen
- Hess, R.D., 1965. Inventory of compensatory education projects. Univ. of Chicago School of Educ.
- Hilgard, E.R., 1959. Creativity and Problem-solving. In: Creativity and its Cultivation (Hrsg. H.H. Anderson). New York, Evanston, 162-180
- Hiltmann, H., 1969. Kompendium der psychodiagnostischen Tests. 2. Aufl., Bern, Stuttgart
- Hitpaß, J., 1961. Vergleichende Untersuchung über den Voraussagewert von Aufnahmeprüfung und Testprüfung zur Erfassung der Eignung für weiterführende Schulen. Schule u. Psychol., 8, 65-71
- Hofstatter, P.R., 1954. The changing composition of "intelligence". A study in T-Technique. J. Genet. Psych., 85, 159-164
- Hofstatter, P.R., 1957. Psychologie. Frankfurt/M.
- Hofstatter, P.R. u. Wendt, D., 1966. Quantitative Methoden der Psychologie. 2. Aufl., München

- Höfler, R., 1927. Über die Bedeutung der Abstraktion für die geistige Entwicklung des taubstummen Kindes. Z. f. Kinderforsch., 33, 414-444
- Höfler, R., 1934. Vergleichende Intelligenzuntersuchungen bei Hörenden und Tauben mit stummen Tests. 4. Vers. Ber. Dtsch. Ges. f. Stimm- und Sprachheilkunde. Leipzig (Sonderdruck)
- Höfler, R., 1952. Die Schwachbegabten. Neue Blätter f. Taubstummenbildg., 6, 325-338
- Hörmann, H., 1961. Zur Validierung von Persönlichkeitstests, insbesondere von projektiven Verfahren. Psychol. Rdsch., 12, 44-49
- Hörmann, H., 1967. Psychologie der Sprache. Berlin, Heidelberg, New York
- Hörr, G.; Keck-Reinhard, G.; Winkler, E.; Roth, J.; 1972. Zur Profilanalyse des P-S-B bei lernbehinderten Sonderschülern. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Holfeld, W., 1968. Soziokulturelles Milieu und Begabung – Intelligenzdiagnostische und statistische Untersuchungen hörgeschädigter Schüler. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Holland, J.L., 1964. The Assessment and Prediction of Creative Performance of High-Aptitude Youth. In: Widening Horizons in Creativity (Hrsg. C.W. Taylor). New York etc., 298-315
- Hollstein, W., 1953. Sprache und Intelligenz in ihrer Wechselwirkung beim taubstummen und hörenden Kind. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Hopf, D., 1971. Entwicklung der Intelligenz und Reform des Bildungswesens. Neue Sammlung, 11, 33-51
- Horn, H., 1958. Psychologische Untersuchungsverfahren während der Früherziehung blinder und hochgradig sehbehinderter Kinder. Der Blindenfreund, 88, 90-101
- Horn, H., 1970. Psychologische Begutachtung und Beurteilung Blinder und hochgradig Sehbehinderter. In: 26. Dt. Blindenlehrerkongr. 1969 in München (Kongreßbericht). Hannover, 156-162
- Horn, H., u. Schwarz, E., 1967. Bildertest BT 1-2 (Intelligenztest für 1. und 2. Klassen). Dt. Bearbeitung d. Moray House Picture Intelligence Tests 1 von M.A. Mellone u. G.H. Thomson. Weinheim, Berlin, Basel
- Horn, W., 1956. Begabungstestsystem (BTS). Göttingen
- Horn, W., 1962. Leistungsprüfsystem (LPS). Göttingen
- Horn, W., 1969. Prüfsystem für Schul- und Bildungsberatung (PSB). Göttingen
- Huber, J.T., 1961. Report writing in psychology and psychiatry. New York
- Hudelmayer, D., 1970. Nichtsprachliches Lernen von Begriffen. Stuttgart
- Hunt, J. McV., 1961. Intelligence and experience. New York
- Hunt, J. McV., 1964. The psychological basis for using pre-school enrichment as an antipode for cultural deprivation. Merrill-Palmer Quart., 10, 209-248
- Hylla, E., 1927. Testprüfungen der Intelligenz. Braunschweig, Berlin, Hamburg
- Hylla, E. u. Kraak, B. (Neubearbeitung von H. Horn, E. Schwarz und U. Raatz), 1965. (1970). Aufgaben zum Nachdenken AzN 4+. Begabungstest für den Übergang auf weiterführende Schulen. Weinheim, Berlin
- Ingenkamp, K., 1963. Psychologische Tests für die Hand des Lehrers. Weinheim
- Ingenkamp, K., 1966. Bildertest BT 2-3 (Intelligenztest für 2. und 3. Klassen). Dt. Bearbeitung d. Deeside Picture Tests von W.G. Emmet. Weinheim, Berlin, Basel
- Ingenkamp, K., 1969. Möglichkeiten und Grenzen des Lehrerurteils und der Schultests. In: Begabung und Lernen (Hrsg. H. Roth). 3. Aufl., Stuttgart, 407-431
- Ingenkamp, K. (Hrsg.), 1971. Tests in der Schulpraxis. Weinheim, Berlin, Basel
- Ingenkamp, K. u. Marsolek, T. (Hrsg.), 1968. Möglichkeiten und Grenzen der Testanwendung in der Schule. Weinheim, Berlin, Basel
- Irle, M., 1955. Berufs-Interessen-Test (B-I-T). Göttingen
- Jäger, A.O., 1967. Dimensionen der Intelligenz. Göttingen
- Janke, W., 1964. Klassifikation. In: Hb. d. Psychol., Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß). Göttingen, 901-929

- Jaspén, N., 1946. Serial correlation. *Psychometrika*, 11, 23-30
- Jensen, A.R., 1967. Social class determinants of language development. In: *Psychology of Language, Thought and Instruction* (Hrsg. J.P. De Cecco). New York, 103-107
- Jones, W.R., 1960. The Development of a non-verbal Tactual Intelligence Scale for the Adult Blind. Diss. Purdue University Ann Arbor, Michigan (USA)
- Jordan, J.E. u. Felty, J., 1968. Factors Associated with Intellectual Variation Among Visually Impaired Children. In: *Research Bulletin* (Hrsg. L.L. Clarke), Nr. 15, 61-70
- Just, G. (Hrsg.), 1939. *Handbuch der Erbbiologie des Menschen*. Band V, 1. Teil. Berlin
- Juurmaa, J., 1967. Ability structure and loss of vision. New York
- Kagan, J. et al., 1958. Personality and IQ change. *J. abnorm. soc. Psychol.*, 56, 261-266
- Keil, W. u. Keil-Specht, H., 1970. Leistungsmotivation und Erziehungsstile. Eine Familienuntersuchung. *Z. f. Entwicklungspsychol. u. Päd. Psychol.*, 2, 241-256
- Kemmler, L., 1967. Erfolg und Versagen in der Grundschule. Göttingen
- Kemmler, L., u. Heckhausen, H., 1967. Ist die sogenannte "Schulreife" ein Reifungsproblem? In: *Pädagogische Psychologie* (Hrsg. F. Weinert). Köln, Berlin, 487-513
- Kerekjarto, M.v. u. Schmidt, G., 1962. Faktorenanalysen des Hamburg-Wechsler-Intelligenztests für Kinder (HAWIK). *Diagnostica*, 8, 95-110
- Kern, A. (Hrsg.), 1970. Die Schulreife in pädagogischer und psychologischer Sicht. Frankfurt/M.
- Kern, A. u. Kern, E., 1969. *Praxis des ganzheitlichen Lesenlernens*. 14. Aufl., Freiburg/Br.
- Kern, E., 1958. Theorie und Praxis eines ganzheitlichen Sprachunterrichts für das gehörgeschädigte Kind. Freiburg/Br.
- Kerstan, G., 1957. Das logische Schließen beim Gehörlosen. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Kettel, K.J. u. Simmat, W.E., 1968. Geschlecht, Alter und Bildung als Bedingungen der Interessenausprägung. *Diagnostica*, 14, 156-173
- Kirk, S.A., 1958. Early Education of the Mentally retarded. *Urbana/III.* – Deutsch 1964: Die Erziehung des zurückgebliebenen Kindes. München, Basel
- Kirk, S.A., 1962. *Educating Exceptional Children*. Boston
- Kirk, S.A.; McCarthy, J.J.; Kirk, W.D., 1968. *The Illinois Test of Psycholinguistic Abilities*. 2. Aufl., Urbana/III.
- Klasen, E., 1971. *Das Syndrom der Legasthenie*. 2. Aufl., Bern, Stuttgart, Wien
- Klauer, K.J., 1962. Sehschwäche und Intelligenz. HAWIK-Untersuchungen bei sehschwachen Kindern. *Z. exp. angew. Psychol.*, 9, 570-593
- Klauer, K.J., 1964. Der Progressive-Matrices-Test bei Volks- und Hilfsschulkindern. *Heilpäd. Forschg.*, 1, 13-37
- Klauer, K.J., 1967. Über den Effekt eines Schulreifetrainings für die Behandlung der Intelligenzschwäche. In: *Pädagogische Psychologie* (Hrsg. F. Weinert). Köln, Berlin, 431-439
- Klauer, K.J., 1969. *Lernen und Intelligenz*. Weinheim, Berlin, Basel
- Klinghammer, H.D., 1968. Verfahren für eine Frühdiagnose. In: *Früherziehung für hörgeschädigte Kinder* (= Sonderheft 1 d. *Z. f. hörgeschädigte Kinder*). Kettwig/Ruhr, 17-23
- Klinghammer, H.D., 1969. Systematische Verhaltensbeobachtung als diagnostisches Hilfsmittel bei sprachentwicklungsgestörten Kindern. *Heilpäd. Forschg.*, 2, 84-94
- Klix, F.; Gutjahr, W.; Mehl, J. (Hrsg.), 1967. *Intelligenzdiagnostik. Probleme und Ergebnisse intelligenzdiagnostischer Forschungen in der DDR*. VEB, Berlin
- Klopfer, W.G., 1960. *The psychological report*. New York
- Köhn, B., 1970. Wortschatzprüfung bei Blinden. Eine empirische Untersuchung mit den Frankfurter Wortschatztests WST 5-6 und 7-8. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Kopfmann, M., 1967. Vergleichende Intelligenzuntersuchungen an hörgeschädigten Kindern: Progressive Matrices – HAWIK. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Kornmann, R., 1971. *Hirnschädigung und fehlende Schulreife*. Berlin

- Kornmann, R., 1972. Untersuchungen über Möglichkeiten der Früherfassung lernbehinderter Sonderschüler mit Hilfe objektiver Testverfahren. Z. f. Entwicklungspsychol. u. Päd. Psychol., 4, 17-26
- Kornmann, R.; Endrigkeit, F.; Sander, H., 1972. Sind lernbehinderte Sonderschüler in Gruppen-Intelligenztests benachteiligt? Zwei vergleichend-empirische Untersuchungen. Diagnostica, 18, 111-121
- Kossakowski, A. u. Lompscher, J. (Hrsg.), 1971. Ideologisch-theoretische und methodologische Probleme der Pädagogischen Psychologie. VEB, Berlin
- Kratzmeier, H.; 1970. Sprach- und Denktraining – Werkspielblätter für 4- bis 7jährige. Weinheim
- Krenberger, S., 1927. Intelligenzprüfung taubstummer Schüler. München
- Kroh, O., 1939. Erbpsychologie der Berufsneigung und der Berufseignung sowie der Sonderbegabungen. In: Hb. d. Erbbiologie d. Menschen (Hrsg. G. Just), Bd. V, 1. Teil. Berlin, 592-668
- Lang, G., 1967. Vergleichende Intelligenzuntersuchungen an hörgeschädigten Kindern: Progressive Matrices – Snijders-Oomen-Test. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Latscha, F., 1966. Der Einfluß des Primarlehrers. In: Die Ungleichheit der Bildungschancen (Hrsg. H. Popitz). Freiburg, Olten, 183-258
- Lawton, D., 1968. Social class, language, and education. London. – Deutsch 1970: Soziale Klasse, Sprache und Erziehung. Düsseldorf
- Leiter, R.G., 1969. The Leiter International Performance Scale. Chicago, Ill.
- Lenneberg, E.H., 1967. Biological foundations of language. New York
- Lewis, M.M., 1963. Language, Thought and Personality in Infancy and Childhood. London. – Deutsch 1970: Sprache, Denken und Persönlichkeit im Kindesalter. Düsseldorf
- Lienert, G.A. u. Fickert, H., 1958. Analyse des verbalen Teils des HAWIE. Marburg (unveröffentl. Manuskript.)
- Lienert, G.A. u. Faber, C., 1963. Über die Faktorenstruktur des HAWIK auf verschiedenen Alters- und Intelligenzniveaus. Diagnostica, 9, 3-11
- Lienert, G.A., 1964. Mechanisch-Technischer-Verständnis-Test (MTVT). Göttingen
- Lienert, G.A., 1967. Testaufbau und Testanalyse. 2. Aufl., Weinheim, Berlin
- Lindner, R., 1925. Vergleichende Intelligenzprüfungen. In: Päd.-Psychol. Arbeiten aus dem Inst. d. Leipziger Lehrervereins (Hrsg. M. Döring), Bd. 14: Das taubstumme Kind in Vergleich mit vollsinnigen Schulkindern. Leipzig, 67-208
- Löwe, A., 1962. Haus-Spracherziehung für hörgeschädigte Kleinkinder. 2. Aufl., 1965, Berlin
- Löwe, A., 1964. Sprachfördernde Spiele für hörgeschädigte Kleinkinder. 2. Aufl. 1967, Berlin
- Löwe, A. u. Heller, K., 1972. Heidelberger Hörprüf-Bild-Test (HHBT) für Schulanfänger. Villingen
- Lückert, H.R. (Hrsg.), 1969 a. Begabungsforschung und Bildungsförderung als Gegenwartsaufgabe. München, Basel
- Lückert, H.R., 1969 b. Die Stabilität und Veränderung der kognitiven Leistungen – Eine Zusammenfassung der wichtigsten Ergebnisse von Benjamin S. Bloom. In: Begabungsforschung und Bildungsförderung (Hrsg. H.R. Lückert). München, Basel, 145-167
- Lückert, H.R., 1969 c. Die basale Begabungs- und Bildungsförderung. In: Begabungsforschung und Bildungsförderung (Hrsg. H.R. Lückert). München, Basel, 225-279
- Lüer, G., 1967. Multivariate Analysen des Testverhaltens familiär schwachsinniger, exogen schwachsinniger und mongoloider Kinder. In: Ber. 25. Kongr. Dt. Ges. Psychol. (Hrsg. F. Merz). Göttingen, 469-474
- Mansholt, E., 1961. Bericht über Versuche mit dem Berliner Begabungstest B 1 in fünf westdeutschen Blindenschulen. Der Blindenfreund, 81, 19-28
- Marsolek, T. u. Ingenkamp, K., 1968. Literatur über Tests im Bereich der Schule. Weinheim, Berlin, Basel

- Martin, E., 1961. Experimentalpsychologische Untersuchungen zur Entwicklung der Konzentrationsfähigkeit bei gehörlosen Schulkindern. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Massialas, B.E. u. Zevin, J., 1969. Kreativität im Unterricht. Unterrichtsbeispiele nach amerikanischen Lerntheorien. Stuttgart
- McClelland, D.C. (Hrsg.), 1955. Studies in motivation. New York
- McClelland, D.C., 1961. The Achieving Society. Princeton, N.J. – Deutsch 1966: Die Leistungsgesellschaft. Stuttgart
- McClelland, D.C.; Atkinson, S.; Clarke, R.; Lowell, E., 1953. The achievement motive. New York
- Meili, R., 1944. Grundlegende Eigenschaften der Intelligenz. Schweiz. Z. f. Psychol., 2, 166-175, 265-271
- Meili, R., 1961. Lehrbuch der psychologischen Diagnostik. 4. Aufl., Bern, Stuttgart
- Meili, R., 1964. Die faktorenanalytische Interpretation der Intelligenz. Schweiz. Z. f. Psychol., 23, 135-155
- Michel, L., 1964. Allgemeine Grundlagen psychometrischer Tests. In: Hb. d. Psychol., Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß). Göttingen, 19-70
- Mierke, K., 1963. Begabung, Bildung und Bildsamkeit. Bern, Stuttgart
- Mittenecker, E., 1964. Subjektive Tests zur Messung der Persönlichkeit. In: Hb. d. Psychol. Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß). Göttingen, 461-487
- Mittenecker, E. u. Toman, W., 1951. Der PI-Test (P-I-T). Ein kombinierter Persönlichkeits- und Interessentest. Wien
- MMPI Saarbrücken, 1963. Deutsche Bearbeitung des Minnesota Multiphasic Personality Inventory von S.R. Hathaway u. J.C. McKinley. Bern, Stuttgart, Wien
- Mühle, G., 1969. Definitions- und Methodenprobleme der Begabtenforschung. In: Begabung und Lernen (Hrsg. H. Roth). 3. Aufl., Stuttgart, 69-97
- Mühle, G. u. Schell, C. (Hrsg.), 1970. Kreativität und Schule – Texte. München
- Neumann, H., 1968. Sprachliche Einschläge in sogenannten sprachfreien Tests. Kettwig/Ruhr
- Newman, H.H.; Freeman, F.N.; Holzinger, K.J., 1937. Twins: A study of heredity and environment. Chicago
- Newman, J.R. u. Loos, F.M., 1955. Differences between verbal and performance IQ's with mentally defective children on the Wechsler intelligence scale for children. J. consult. Psychol., 19, 16 ff.
- Nickel, H., 1967. Die visuelle Wahrnehmung im Kindergarten- und Einschulungsalter. Bern, Stuttgart
- Nickel, H., 1969. Die Bedeutung planmäßiger Übung für die Entwicklung einer differenzierenden visuellen Auffassung im Vorschulalter. Z. f. Entwicklungspsychol. u. Päd. Psychol., 1, 103-118
- Niepold, W., 1971. Sprache und soziale Schicht. 2. Aufl., Berlin
- Oevermann, U., 1967. Sprache und soziale Herkunft – Ein Beitrag zur Analyse schichtenspezifischer Sozialisationsprozesse und ihrer Bedeutung für den Schulerfolg. Diss. Frankfurt/M.
- Oevermann, U., 1969. Schichtenspezifische Formen des Sprachverhaltens und ihr Einfluß auf die kognitiven Prozesse. In: Begabung und Lernen (Hrsg. H. Roth). 3. Aufl., Stuttgart, 297-356
- Oléron, P. u. Herren, H., 1961. L'acquisition des conversations et le langage: Etudes comparatives sur des enfants sourds et entendants. Enfance, 14, 203-219
- Orlik, P., 1967. Kritische Untersuchungen zur Begabtenförderung. Meisenheim am Glan
- Osgood, C.E., 1957. A behavioristic analysis of perception and language as cognitive phenomena. In: Contemporary approaches to cognition (Hrsg. J. S. Bruner et al.). Cambridge, Mass., 75-118
- Pawlik, K., 1968. Dimensionen des Verhaltens. Bern, Stuttgart
- Pawlik, K., 1970. Psychologische Diagnostik. Hamburg (Vorlesungsskript)



- Pongratz, L., 1960. Frühkindliche Prägung und Charakterentwicklung. Jb. f. Psychol. u. Psychotherapie, 7, 312-324
- Priester, H.J., 1958. Die Standardisierung des Hamburg-Wechsler-Intelligenztests für Kinder (HAWIK). Bern, Stuttgart
- Priester, H.J., 1959. Untersuchungen zum Vergleich zwischen Stanford-Intelligenztest (SIT) und Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK). Diagnostica, 5, 49-58
- Priester, H.J., 1964. Intelligenztests für Erwachsene. In: Hb. d. Psychol., Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß). Göttingen, 226-259
- Priester, H.J.; Kerekjarto, M.v., 1960. Weitere Forschungsergebnisse zum HAWIE und HAWIK. Diagnostica, 6, 86-94
- Raatz, U., 1971. Der Einsatz von Tests bei der Kontrolle von Schulversuchen. In: Tests in der Schulpraxis (Hrsg. K. Ingenkamp). Weinheim, 172-182
- Rath, W., 1967. Untersuchungen bei Blinden und hochgradig Sehbehinderten zum Vergleich zwischen HAWIK-Verbalteil und I-P-T. Der Blindenfreund, 87, 113-129
- Raven, J.C., 1936. Standard Progressive Matrices (PMT). Revision 1956. London
- Raven, J.C., 1947. Coloured Progressive Matrices. Revision 1956. London
- Raven, J.C., 1958. Advanced Progressive Matrices. 2. Aufl., London
- Reinartz, A. u. Seifart, O., 1970. Profilauswertung des Hamburg-Wechsler-Intelligenztests für Kinder (HAWIK). 2. Aufl., Berlin
- Rexhausen, J., 1964. Versuch zur Eichung des Intelligenz-Punkt-Tests nach Drummond-Strehle. Der Blindenfreund, 84, 170-186
- Rich, C.C., 1963. The Validity of an Adaptation of Raven's Progressive Matrices Test for Use with Blind Children. Diss. am Texas Technological College
- Riegel, R.M., 1960. Faktorenanalysen des HAWIE für die Altersstufen 20-34, 35-49, 50-64 und 65 Jahre und älter. Diagnostica, 6, 41-66
- Ritter, H., u. Engel, W., 1969. Genetik und Begabung. In: Begabung und Lernen (Hrsg. H. Roth). 3. Aufl., Stuttgart, 99-128
- Rösler, H.D.; Schmidt, H.D.; Szweczyk, H. (Hrsg.), 1970. Persönlichkeitsdiagnostik. Probleme und Ergebnisse persönlichkeitsdiagnostischer Forschungen in der klinischen Psychologie der DDR. VEB, Berlin
- Rosen, B.C., 1956. The achievement syndrome. Am. Sociol. Review, 21, 203-212
- Roth, E.; Oswald, W.D.; Daumenlang, K., 1972. Intelligenz. Urban-Taschenb., Bd. 144. Stuttgart usw.
- Roth, H. (Hrsg.), 1969. Begabung und Lernen. (= Bd. 4 der Gutachten und Studien der Bildungskommission des Deutschen Bildungsrates). 3. Aufl., Stuttgart
- Sader, M., 1957. Instruktionsverständnis und Testleistung. Frankfurt/M.
- Sader, M. u. Keil, W., 1966. Bedingungskonstanz in der psychologischen Diagnostik (Sammelreferat). Arch. f. d. ges. Psychol., 118, 279-308
- Samstag, K. u. Baus, M., 1962. Pädagogisch-psychologisches Testen – Anleitung zur Testarbeit in der Schule. Bad Heilbrunn/Obb.
- Sapir, E., 1961. Die Sprache. München
- Sarason, S.B., 1957. Test anxiety, general anxiety, and intellectual performance. J. consult. Psychol., 21, 485-490
- Sarason, S.B. et al., 1958. A test anxiety scale for children. Child Developmt., 29, 105-113
- Sayler, W., 1965. Die Interessenstruktur von Studierenden an Pädagogischen Hochschulen. Schule u. Psychol., 12, 203-210
- Scheler, M., 1962. Die Stellung des Menschen im Kosmos. 6. Aufl., München
- Schirmer, B., 1971. Wortschatzprüfung bei Sehbehinderten. Eine empirische Untersuchung mit den Frankfurter Wortschatztests WST 5-6 und WST 7-8. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Schmalohr, E., 1963. Die Profilanalyse des HAWIK in der Hilfsschulauslese. Z. f. Heilpäd., 14, 12-24

- Schmalohr, E., 1971 a. HAWIK-Zusatzauswertung. 3. Aufl., Krefeld
- Schmalohr, E., 1971 b. Psychologie des Erstlese- und Schreibunterrichts. 2. Aufl., München
- Schmidt, J., 1958. Erste Grundlegung zur Entwicklung psychologischer Untersuchungsverfahren an Blinden. Z. exp. angew. Psychol., 5, 295-338
- Schmidt, J., 1960. Das Problem psychologischer Eignungsuntersuchungen für Blinde. Der Blindenfreund, 80, 131-143
- Schollmeyer, B., 1970. Erfahrungen mit dem "Fremdspracheneignungstest für die Unterstufe" an der Staatlichen Schwerhörigenschule in Waldkirch. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Schraml, W., 1964. Das Psychodiagnostische Gespräch (Exploration und Anamnese). In: Psychol. Diagnostik, Bd. 6 d. Hb. d. Psychol. (Hrsg. R. Heiß). Göttingen, 868-897
- Schubenz, S., 1967. Über den Aufbau eines psychologischen Befundes. Diagnostica, 13, 89-116
- Schubenz, S. u. Buchwaldt, R., 1964. Untersuchungen zur Legasthenie. Z. f. exp. angew. Psychol., 11, 155-168
- Schulte, K., 1968. Der Sinnbezirk – Gegenstand der Wortinhaltsforschung und Voraussetzung des Sprachaufbaus bei Hörgeschädigten. Kettwig/Ruhr
- Schulte, K., 1969. Wortschatz, Arbeit am . . . In: Enzyklopäd. Hb. d. Sonderpädagogik (Hrsg. G. Heese u. H. Wegener). Berlin, 3815-3819
- Schultze, W. et al., 1964. Über den Voraussagewert der Auslesekriterien für den Schulerfolg an Gymnasien. (= Forschungsbericht I der Max-Traeger-Stiftung). Frankfurt/M.
- Schweingruber, W., 1955. Begabungsforschungen bei Blinden. St. Gallen (Schw.)
- Seifert, K.H., 1960 a. Der Progressive-Matrices-Test und seine Anwendung bei taubstummen Kindern. Neue Blätter f. Taubstummenbild., 14, 16-22
- Seifert, K.H., 1960 b. Die Problematik der Intelligenzforschung bei taubstummen Kindern. Neue Blätter f. Taubstummenbild., 14, 240-250
- Seifert, K.H., 1960 c. Erfahrungen mit dem Progressive-Matrices-Test bei taubstummen Kindern. Z. f. exp. angew. Psychol., 7, 255-290
- Shurrager, H.S. u. Shurrager, P.S., 1964. Haptic Intelligence Scale for Adult Blind (HISAB). Chicago, Illinois (USA)
- Skowronek, H., 1968. Psychologische Grundlagen einer Didaktik der Denkerziehung. Hannover usw.
- Snijders, J.T. u. Snijders-Oomen, N., 1958 a. Nichtverbale Intelligenztestreihe (SON). Groningen
- Snijders, J.T. u. Snijders-Oomen, N., 1958 b. Sprachfreie Intelligenzuntersuchung für Hörende und Taubstumme – Textband zum SON. Groningen
- Solomon, P. et al., 1961. Sensory deprivation. Cambridge/Mass.
- Sommer, G., 1971. Diagnostische Psychologie. In: Steckbrief der Psychologie (Hrsg. K.E. Rogge). Heidelberg, 190-206
- Spearman, C., 1904. General intelligence objectively determined and measured. Am. J. Psychol., 15, 201-292
- Spearman, C., 1927. The abilities of man. London
- Spitzmüller, R., 1969. Die sozialen Determinanten der Begabungsentwicklung und des Leistungsstrebens. In: Begabungsforschung und Bildungsförderung als Gegenwartsaufgabe (Hrsg. H.R. Lückert). München, Basel, 184-195
- Spitznagel, A., 1964. Die diagnostische Situation. Ein Beitrag zur Theorie und Psychologie der Datengewinnung. Habil.Schrift Freiburg/Br.
- Steinkamp, G., 1967. Die Rolle des Volksschullehrers im schulischen Selektionsprozeß. In: Hamburger Jb. f. Wirtschafts- und Sozialpolitik, 12. Tübingen
- Stern, W., 1911. Die differentielle Psychologie in ihren methodischen Grundlagen. Leipzig
- Stern, W., 1912. Die psychologischen Methoden der Intelligenzprüfung und deren Anwendung an Schulkindern. Leipzig (5. Kongr. Exp. Psychol., Berlin 1912)
- Stern, W., 1950. Allgemeine Psychologie auf personalistischer Grundlage. 2. Aufl., Den Haag

- Strauß, M., 1967. Versuch einer Anwendung der Vocational Intelligence Scale for Adult Blind (VISAB) bei 10- bis 15jährigen Kindern. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Strehle, W., 1961 a. Ein Binet-Test für Blinde. Der Blindenfreund, 81, 105-129
- Strehle, W., 1961 b. Ein Intelligenz-Punkt-Test für blinde Kinder und Jugendliche. Der Blindenfreund, 81, 159-171, 178-186
- Süllwold, F., 1954. Experimentelle Untersuchungen über die Rolle des Einfalles im Denkprozeß. Z. f. exp. angew. Psychol., 2, 175-207
- Süllwold, F., 1959. Bedingungen und Gesetzmäßigkeiten des Problemlösungsverhaltens. Ber. Kongr. Dtsch. Ges. Psychol., 22, 96-115
- Süllwold, F., 1963. Einführung in die quantitativen Methoden der Psychologie II. Univ. Heidelberg (Vorl.-Mitschr.)
- Süllwold, F., 1964. Schultests. In: Hb. d. Psychol., Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß). Göttingen, 352-384
- Taylor, C.W. (Hrsg.), 1964. Widening Horizons in Creativity. New York etc.
- Templin, M.C., 1950. The development of reasoning in children with normal and defective hearing. In: Inst. Child Welfare Monogr., Ser. Nr. 24. Minneapolis: Univ. Minnesota Press.
- Tent, L., 1969. Die Auslese von Schülern für weiterführende Schulen. Göttingen
- Thomae, H., 1967. Prinzipien und Formen der Gestaltung psychologischer Gutachten. In: Hb. d. Psychol., Bd. 11: Forensische Psychologie (Hrsg. U. Undeutsch). Göttingen 743-767
- Thomae, H., 1970. Beobachtung und Beurteilung von Kindern und Jugendlichen. 9. Aufl. Basel (Schw.), New York
- Thurner, F. u. Tewes, U., 1969. Der Kinder-Angst-Test (KAT). Göttingen
- Thurstone, L.L., 1938. Primary mental abilities. Chicago, Psychometric Monogr., 1
- Thurstone, L.L., 1945. Theories of intelligence. Chicago
- Tillman, M.H., 1967. The performances of blind and sighted children on the Wechsler Intelligence Scale for Children. Intern. J. for the Educ. of the Blind, 16, 65-74, 106-112
- Überla, K., 1968. Faktorenanalyse – Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler. Berlin, Heidelberg, New York
- Ulmann, G., 1968. Kreativität – Neue amerikanische Ansätze zur Erweiterung des Intelligenzkonzeptes. Weinheim, Berlin, Basel
- Undeutsch, U., 1969. Zum Problem der begabungsgerechten Auslese beim Eintritt in die höhere Schule und während der Schulzeit. In: Begabung und Lernen (Hrsg. H. Roth). Stuttgart, 377-405
- Vernon, P.E., 1961. The structure of human abilities. London
- Vukovich, A.F., 1964. Die Konstruktion psychologischer Tests. In: Hb. d. Psychol., Bd. 6: Psychologische Diagnostik (Hrsg. R. Heiß). Göttingen, 113-138
- Wappmann, H., 1963. Testuntersuchung als Beitrag zur Erforschung des logischen Denkens blinder und sehbehinderter Kinder. Examensarbeit am Sonderpäd. Inst. Heidelberg (unveröffentl.)
- Weber, E., 1967. Grundriß der biologischen Statistik. 6. Aufl., Stuttgart
- Wechsler, D., 1961. Die Messung der Intelligenz Erwachsener. 2. Aufl., Bern, Stuttgart. (Siehe auch C. Bondy 1956)
- Weinert, F., 1964. Experimentelle Untersuchungen über Formen und Bedingungen des kognitiven Lernens. Arch. ges. Psychol., 116, 126-164
- Weinert, F. (Hrsg.), 1967. Pädagogische Psychologie. 7. Aufl. 1972, Köln, Berlin
- Weiß, R., 1969. Die Brauchbarkeit des Culture Free Intelligence Tests Skala 3 (CFT 3) bei begabungspsychologischen Untersuchungen. Diss. Würzburg
- Wendeler, J., 1967. Untersuchung von Zusammenhängen zwischen Persönlichkeitsmaßen und Intelligenzleistungen. In: Psychologie und Pädagogik (Hrsg. H. Horn). Weinheim, 107-117
- Wendeler, J., 1969. Standardarbeiten – Verfahren zur Objektivierung der Notengebung. Weinheim, Berlin, Basel

- Wendeler, J., 1970. Intelligenztests in Schulen. Weinheim, Berlin, Basel
- Wenzl. A., 1957. Theorie der Begabung. 2. Aufl., Heidelberg
- Wertheimer, M., 1945. Productive thinking. 2. Aufl. 1959, New York. – Deutsch: Produktives Denken (Hrsg. W. Metzger). Frankfurt/M.
- Wewetzer, K.H., 1964. Intelligenztests für Kinder. In: Hb. d. Psychol., Bd. 6: Psychol. Diagnostik (Hrsg. R. Heiß). Göttingen 200-225
- Wewetzer, K.H., 1972. Intelligenz und Intelligenzmessung. Darmstadt
- Whorf, B.L., 1963. Sprache – Denken – Wirklichkeit. Hamburg (rde 174)
- Wilde, K., 1951. Über die Zuverlässigkeit psychologischer Untersuchungsmethoden. Psychol. Rdsch., 2, 187-193
- Wittmann, J., 1933. Theorie und Praxis eines ganzheitlichen analytisch-synthetischen Unterrichts. Potsdam
- Witzlack, G., 1971. Methodologische Probleme der pädagogisch-psychologischen Diagnostik In: Ideologisch-theoretische und methodologische Probleme der Pädagogischen Psychologie (Hrsg. A. Kossakowski u. J. Lompscher). Berlin, 125-148
- Wolf, R.M., 1963. The identification and measurement of environmental process variables related to intelligence. Ph. D. Diss. Univ. of Chicago
- Wooden, H.Z., 1963. Deaf and hearing children. In: Exceptional Children in the Schools (Hrsg. L.M. Dunn). New York, 339-411
- Zeckel, A. u. Kolk, v.d., 1939. Eine vergleichende Intelligenz-Untersuchung einer Gruppe erblich taubstummer und hörender Kinder mittels der Poteus-Intelligenz-Probe. Rotterdam
- Ziehen, T., 1933. Intelligenz und Sprachentwicklung. Blätter f. Taubstummenbildg., 46, 113-120
- Zimmermann, K.W., 1969. Psychodiagnostische Verfahren zur Untersuchung von Lernbehinderten. Berlin
- Zimmermann, K.W. u. Kornmann, R., 1970. Die Profilanalyse des HAWIK bei lernbehinderten Sonderschülern. Z. f. Heilpäd., 21, 555-563
- Zimmermann, K.W.; Kornmann, R.; Lorenz, A.L., 1971. Der HAWIK bei lernbehinderten Sonderschülern. Oberbiel/Wetzlar

**Anm.:** Die Publikationen von Allinger u. Heller (1972) sowie Heller (1972 c) erscheinen Anfang 1973; entsprechende Seitenverweise im Text beziehen sich auf das Manuskript.

## Verzeichnis der Grafiken

- Grafik 1    Standardabweichungen ( $\sigma$ ) als Parameter der normierten Gaußschen Verteilung sowie Akzeptationsbereiche (Wahrscheinlichkeiten in %) versus Irrtums-(Rest-)Wahrscheinlichkeiten für verschiedene Signifikanzniveaus – jeweils als Flächenanteile unter der Normalkurve
- Grafik 2    Transformation einer Rohwertskala (Rohpunkte) in %-Ränge (PR) und Standardwertskalen (z-Werte, T-Werte, Centil-Werte, Abweichungs-IQ) sowie Flächen-Prozente (% der Fälle über den Achsenabschnitten der Standardabweichungen) unter der Normalkurve
- Grafik 3    Stabdiagramme der empirischen Testleistungswerte im WST, PSB, AzN 4+ und CFT des 4. Grundschulklassenkollektivs einer süddeutschen Großstadt (N = 3534)
- Grafik 4    Stabdiagramme der empirischen Testleistungswerte im WST, PSB, AzN 4+ und CFT dreier Schulstichproben des 5. Schuljahrs einer süddeutschen Großstadt
- Grafik 5    Eignungskennbereiche im LPS für *Hauptschul*begabungen (H-Klassen 6 und 8)
- Grafik 6    Eignungskennbereiche im LPS für *Realschul*begabungen (R-Klassen 6 bis 10)
- Grafik 7    Eignungskennbereiche im LPS für *Gymnasial*begabungen (G-Klassen 6 bis 12)
- Grafik 8    Eignungskennbereiche im PSB für Hauptschüler, Realschüler und Gymnasialisten (empirische Gruppenwerte der Schj./Kl. 5)
- Grafik 9    Summenprozentverteilung des HAWIK-Gesamt-IQ bei Grundschülern und Sonderschulanwärtern – Klassifikationsschema nach Schmalohr (1971 a)
- Grafik 10   Schema zur HAWIK-Profilanalyse lernbehinderter Sonderschüler nach Reinartz & Seifart (1970)
- Grafik 11   Kumulative Frequenzen (Summenprozentverteilung der Rohwerte) der PSB-Gesamtleistung bei 324 Lb-Sonderschülern der 7. bis 9. Schuljahrklasse – zit. nach Hörr et al. (1972)
- Grafik 12   Durchschnittliche HAWIK-Profile sehbehinderter Schüler nach Klauer (1962)
- Grafik 13   HAWIK-Gruppenprofil sehbehinderter Schüler (N = 73) nach Baitinger & Bernd (1970)
- Grafik 14   Gruppenprofil blinder Schüler (N = 62) im HAWIK-Verbalteil nach Hudelmayer (1970)

# Verzeichnis der Tabellen

## A. Text

- Tabelle 1 Muster eines Erhebungs- und Beurteilungsbogens (FB-Kurzfassung) für die Bildungsberatung, hier Schullaufbahnberatung im 4. Grundschuljahr
- Tabelle 2 Auszugsweise Wiedergabe des "Anamnese- und Protokollbogens für sehgeschädigte Kinder" nach H. Horn (1968)
- Tabelle 3 Matrix der Profildifferenzen im LPS (C-Werte) auf der Grundschulstufe (Klasse 4) – Sicherung auf dem 5%-Niveau
- Tabelle 4 Matrix der Profildifferenzen im LPS (C-Werte) in der Anwendung bei Sekundarstufenschülern und Erwachsenen – Sicherung auf dem 5%-Niveau
- Tabelle 5 Matrix der Profildifferenzen im PSB (C-Werte) auf der Grundschulstufe (Klasse 4) – Sicherung auf dem 5%-Niveau
- Tabelle 6 Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei Sekundarstufenschülern bzw. Erwachsenen – Sicherung auf dem 5%-Niveau
- Tabelle 7 Mittelwerte, Sigtabereiche und Variationsbreite (Extremwerte) in C der nach AUKL ermittelten Schulleistungsgruppen für H, R und G im LPS
- Tabelle 8 Korrelative Zusammenhänge von Englisch-Zensur (Engl.) und verschiedenen Test- bzw. Schulleistungskriterien bei 28 Hauptschülern der 5. Klasse – nach U. Beck (1969)
- Tabelle 9 Arithmetische Mittelwerte (M) und Standardabweichungen (s) der Wertpunkt-Verteilungen in den einzelnen HAWIK-Subtests, berechnet aus den WP-Häufigkeiten bei 824 lernbehinderten Sonderschülern – zit. nach Zimmermann et al. (1971)
- Tabelle 10 Häufigkeitsverteilungen der Rohpunkte in den einzelnen PSB-Subtests bei 212 Lb-Verdachtsfällen des 4. Grundschuljahrs (10. Lbj.) sowie 324 Lb-Sonderschülern der 7. bis 9. Klasse bzw. des 13. bis 15. Lebensjahrs
- Tabelle 11 Rohpunkt-Verteilung der PSB-Gesamtleistung bei 212 Lb-Verdachtsfällen des 4. Grundschuljahrs (10. Lbj.) sowie 324 Lb-Sonderschülern der 7. bis 9. Klasse bzw. des 13. bis 15. Lebensjahrs
- Tabelle 12 PSB-Leistungen (M- und s-Werte in T) verschiedener Lb-Gruppen und statistische Kontrolle der Gesamtgruppen-Differenzen (N = 324 vs. N = 212)
- Tabelle 13 Retest- und Paralleltest-Reliabilitätskoeffizienten ( $r_{tt}$  bzw.  $r_p$ ) des PSB in der Anwendung bei 13- bis 15jährigen lernbehinderten Sonderschülern – zit. nach Hörr et al. (1972)
- Tabelle 14 Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei 13jährigen lernbehinderten Sonderschülern – Sicherung auf dem 5%-Niveau
- Tabelle 15 Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei 14jährigen lernbehinderten Sonderschülern – Sicherung auf dem 5%-Niveau
- Tabelle 16 Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei 15jährigen lernbehinderten Sonderschülern – Sicherung auf dem 5%-Niveau
- Tabelle 17 Matrix der Profildifferenzen im PSB (C-Werte) in der Anwendung bei lernbehinderten Sonderschülern der Oberstufe (Klasse 7-9) – Sicherung auf dem 5%-Niveau
- Tabelle 18 Testleistungen (M-Werte) der Lb-Verdachtsfälle im AzN 4+ und CFT bei verschiedenen Untersuchungssamples des 4. Grundschuljahrs 1967/68
- Tabelle 19 Interkorrelationen zum SON und RD (HAWIK) bei 66 tauben/resthörigen Schülern im Alter von 13 bis 15 Jahren – zit. nach Bäumler et al. (1968)
- Tabelle 20 Ergebnisse zur Gültigkeitsbestimmung des PMT, SON und HAWIK-Handlungsteils in der Anwendung bei Hörgeschädigten

Tabelle 21	Interkorrelationen des FTU 4-6 mit verschiedenen Außenkriterien bei drei schwerhörigen Untersuchungssamples (Gesamt-N = 34) – eigene Berechnungen nach den Untersuchungsdaten von Schollmeyer (1970)
Tabelle 22	Subtestprofile (Mittelwerte in WP) im HAWIK bei 466 Legasthenikern, differenziert nach Geschlecht und Schulstufe – zit. nach Klasen (1971)
Tabelle 23	Mittelwerte (M) und Standardabweichungen (s) des HAWIK-Verbal-IQ, HAWIK-Handlung-IQ und HAWIK-Gesamt-IQ bei verschiedenen Stichproben bzw. Visusgruppen sehbehinderter und blinder Schüler
Tabelle 24	Mittelwerte blinder und hochgradig sehbehinderter Hilfs-, Mittel- und Oberschüler in den Subtests des HAWIK-Verbalteils im Vergleich zu den entsprechenden Kennwerten der HAWIK-Standardisierungsstichproben Sehender – nach Rath (1967)
Tabelle 25	IQ-Verteilung Sehgeschädigter im Vergleich zur Normalklassifizierung im HAWIK – nach Baitinger & Bernd (1970)

## *B. Anhang*

Tabelle	I a	Altersnormen zum Progressiven Matrizen-Test für hörgeschädigte Kinder und Jugendliche
Tabelle	I b	Schulnormen zum Progressiven Matrizen-Test für <i>taube</i> Kinder und Jugendliche
Tabelle	I c	Schulnormen zum Progressiven Matrizen-Test für <i>resthörige/schwerhörige</i> Kinder und Jugendliche
Tabelle	I d	Schulnormen zum Progressiven Matrizen-Test für Hörgeschädigte weiterführender gehobener Bildungseinrichtungen (Realschule bzw. Wirtschaftsschule)
Tabelle	I e	Berufsschulnormen zum Progressiven Matrizen-Test für Hörgeschädigte
Tabelle	II a	Alters-Prozentränge (PR) zum Benton-Test (Zeichenform C und Wahlform G) für taube/resthörige Schüler
Tabelle	II b	Altersnormen zum Benton-Test (Zeichenform C und Wahlform G) sowie Fehlerkategorien der Zeichenform C (%-Anteile) für taube/resthörige Schüler
Tabelle	III	Altersnormen zum HAWIK (Handlungsteil) für taube/resthörige Schüler
Tabelle	IV a	Schul-Prozentränge (PR) im Frankfurter Wortschatztest (WST Form I bzw. II) für taube/resthörige Hauptschüler unter Verwendung des WST als Powertest
Tabelle	IV b	Schul-Prozentränge (PR) im Frankfurter Wortschatztest (WST Form I bzw. II) für schwerhörige Hauptschüler unter Verwendung des WST als Powertest
Tabelle	IV c	Schul-Prozentränge (PR) im Frankfurter Wortschatztest (WST Form I bzw. II) für hörgeschädigte Realschüler und Wirtschaftsschüler – Testbearbeitungszeit: 20 Min.
Tabelle	V a	Altersnormen zum Bourdon-Figuren-Durchstreich-Test für taube Schüler (nach E. Martin 1961)
Tabelle	V b	Altersnormen zum Bourdon-Figuren-Durchstreich-Test für resthörige Schüler (nach E. Martin 1961)
Tabelle	VI a	Altersnormen zum Konzentrations-Verlaufs-Test für taube Schüler (KVT <sub>Taube</sub> )
Tabelle	VI b	Altersnormen zum Konzentrations-Verlaufs-Test für resthörige Schüler (KVT <sub>Resthörige</sub> )

<b>Tabelle</b>	<b>VI c</b>	<b>Alters-Prozentränge (PR) zum Konzentrations-Verlaufs-Test für 561 österreichische Schulkinder einschließlich 314 Internats-Sonderschüler (von H. Seyfried; zit. bei E. Martin 1961) – Zeitvariable</b>
<b>Tabelle</b>	<b>VI d</b>	<b>Alters-Prozentränge (PR) zum Konzentrations-Verlaufs-Test für 1234 österreichische Schulkinder einschließlich 314 Internats-Sonderschüler (von H. Seyfried; zit. bei E. Martin 1961) – Fehlervariable</b>
<b>Tabelle</b>	<b>VII a</b>	<b>Altersnormen zum Pauli-Test für taube Schüler (nach E. Martin 1961)</b>
<b>Tabelle</b>	<b>VII b</b>	<b>Altersnormen zum Pauli-Test für taube Jugendliche (von E. Gruhnwald und E. Ulich; zit. nach W. Arnold 1961)</b>
<b>Tabelle</b>	<b>VII c</b>	<b>Altersnormen zum Pauli-Test für resthörige Schüler (nach E. Martin 1961)</b>
<b>Tabelle</b>	<b>VIII</b>	<b>Transformation einschlägiger Testnormen und Zensuren</b>
<b>Tabelle</b>	<b>IX</b>	<b>T-Transformationstabelle (Umwandlung von Rangplätzen in T-Standardwerte) für kleinere Stichproben (nach F. Tremel 1967)</b>



# Personenregister

- Abels, D. 28, 143f.  
Aebli, H. 34, 48  
Affolter, F. 11  
Allan, J.M. 95  
Allinger, U. 74, 98, 129, 138f., 165, 169f., 185  
Amelang, M. 123, 185  
Amthauer, R. 25, 118, 126  
Anastasi, A. 33, 80, 118, 131, 208  
Anderson, H.H. 22f.  
Andre, K. 11  
Anger, H. 65, 86, 146  
Atkinson, J.W. 35f.  
Arnold, W. 152ff., 243  
Aurin, K. 10, 87f., 131, 135, 138f., 164f., 183, 195  
Baitinger, U. 209f., 212, 215f.  
Bangen, R. 197, 201  
Bargmann, R. 146  
Barron, F. 23  
Bartenwerfer, H. 18, 143f., 150, 152f.  
Bartlett, C.J. 123  
Batt, H. 140f., 197, 202  
Bäumler, G. 153, 197ff.  
Baumeister, A.A. 123  
Baus, M. 80  
Beck, U. 148, 182ff.  
Belser, H. 80  
Bender, L. 140  
Benton, A.L. 140ff., 202, 229f.  
Bereiter, C. 42  
Bergius, R. 16  
Bernd, R. 209f., 212, 215f.  
Bernstein, B. 34f., 37ff., 42f., 46ff., 70, 72  
Binet, A. 10, 32f., 47, 52, 56, 81, 117ff., 125, 134, 189, 207  
Bleidick, U. 86, 88, 99  
Bloom, B.S. 30ff., 50f.  
Bobertag, O. 56, 81, 207  
Bondy, C. 114, 119  
Bourdon, B. 28, 113, 143, 150f., 238  
Bracken, H. von 112, 185  
Brennecke, O. 197, 201  
Bronfenbrenner, U. 36f.  
Brown, R.W. 41  
Brown, W. 41, 64  
Bruner, J.S. 46  
Brunner, P. 135, 197  
Buchwald, R. 108, 205  
Burt, C. 14f., 29  
Büscher, P. 197, 201, 205  
Carroll, J.B. 147  
Cattell, J. McK. 56ff.  
Cattell, R.B. 29, 86, 130ff., 154, 189  
Chauncey, H. 32, 62  
Chomsky, N. 51  
Clark, W.W. 120  
Cohen, R. 140  
Cooley, W.W. 165  
Correll, W. 71, 147  
Cronbach, L.J. 61, 80f.  
Dahl, G. 121  
Dalferth, B. 215  
Daniels, J.C. 220  
Dautermann, W.L. 207  
Deller, E. 11  
Demel, E. 21  
Deutsch, C.P. 41f.  
Deutsch, M. 41f.  
Dobbin, J.E. 32, 62  
Doman, G. 48  
Donat, H. 86  
Drever, G. 57, 196  
Düker, H. 143, 151  
Ebbinghaus, H. 10, 56f.  
Eckrich, W. 147, 197, 202f.  
Eigler, G. 82  
Engel, W. 33  
Estes, B.W. 120, 195  
Evans, L. 134, 197, 199, 201  
Eyferth, K. 120  
Faber, C. 123  
Faber, E. 169  
Feldt, L.S. 72  
Felty, J. 216f.  
Fickert, H. 117f.  
Fischer, A. 10  
Fischer, G.H. 55  
Flechsigg, K.K. 50  
Floud, J. 38  
French, J.W. 137, 147  
Fröhlich, W.D. 57, 69f., 196  
Frohn, W. 11  
Furth, H.G. 45f.  
Gaude, P. 53, 55, 63  
Gaus, A. 34, 47  
Gebauer, E. 11  
Gebauer, T. 129  
Gebhardt, H. 197  
Getzels, J.W. 49  
Gleser, G.C. 61, 81  
Goetzinger, C.P. 45  
Gottschaldt, K. 10, 12f., 15, 33, 247  
Graumann, C.F. 11, 20, 84ff., 208  
Gray, S.W. 48  
Groffmann, K.J. 9f., 56ff.  
Gruhnwald, E. 153, 243

Guilford, J.P. 14, 19ff., 24f., 28ff.,  
205, 247  
Gulliksen, H. 74, 80

Hammond, K.R. 95  
Hardesty, A. 114  
Hardesty, F.P. 119  
Hartmann, H. 61 95f., 109f.  
Hartmann, N. 206  
Hasemann, K. 86  
Hayes, S.P. 207, 213ff.  
Heckhausen, H. 34, 36, 48ff.  
Heider, F. 42  
Heider, G.M. 42  
Heiß, R. 9, 95ff., 156  
Hengstler, M. 214f.  
Herderschêe, D. 197  
Herren, H. 45  
Herrmann, T. 37  
Hess, R.D. 49  
Hieronymus, A.N. 37  
Hilgard, E.R. 22ff.  
Hiltmann, H. 151, 156  
Hitpaß, J. 153  
Hoepfner, R. 14, 21  
Höfler, R. 11, 197  
Hofstätter, P.R. 10, 14, 43, 55, 64, 67  
Höhn, E. 129  
Holfeld, W. 197, 205  
Holland, J.L. 23  
Hollstein, W. 45  
Hopf, D. 32  
Hörmann, H. 38, 43, 51, 75  
Horn, H. 88, 100, 128, 146, 207  
Horn, W. 25, 65, 69, 86f., 135ff., 146,  
157f., 160, 167f., 189  
Hörr, G. 189f., 192f.  
Huber, J.T. 95  
Hudelmayer, D. 45, 54, 209, 211ff., 215f.  
Hunt, J. McV. 33, 49  
Hylla, E. 56, 81, 86, 128, 146, 189

Ingenkamp, K. 32, 63, 71, 75, 80, 82f.,  
147, 156  
Irle, M. 145, 154, 204

Jäger, A.O. 14, 16f., 21, 24ff., 33ff., 247  
Janke, W. 169  
Jaspen, N. 138  
Jensen, A.R. 42f.  
Jones, W.R. 141ff.  
Jordan, J.E. 216f.  
Juurmaa, J. 214

Kagan, J. 36  
Kamratowski, J. 148  
Kemmler, L. 34, 50, 163, 183, 185f.  
Keil, W. 36, 61  
Keil-Specht, H. 36  
Kettel, K.J. 155  
Kerekjarto, M. von 123

Kern, A. 34, 48  
Kern, E. 11, 48  
Kerstan, G. 11  
Kirk, S.A. 33f., 52  
Klasen, E. 108, 205f.  
Klauer, K.J. 34, 135, 185, 192, 208ff.  
Klaus, R.A. 48  
Klinghammer, H.D. 86, 197  
Klix, R. 80  
Klopfer, W.G. 95  
Köhn, B. 146f., 215  
Kolk, v. d. 252  
Kopfmann, M. 197, 201  
Kornmann, R. 123, 185, 187, 196  
Kraak, B. 86, 128, 189  
Kraepelin, E. 58, 152f.  
Kratzmeier, H. 48, 54  
Krenberger, S. 197  
Kretschmer, E. 129  
Kroh, O. 10

Lämmermann, H. 81  
Lang, G. 197, 201  
Latscha, F. 47  
Lauber, H. 114  
Lawton, D. 41f., 45  
Leiter, R.G. 220  
Lenneberg, E.H. 45  
Lersch, P. 10, 13  
Lewis, M.M. 46  
Lichtenberger, W. 135  
Lienert, G.A. 61ff., 69, 71ff., 78, 117f.,  
123ff., 149ff., 157, 159f.  
Lindner, R. 197  
Löwe, A. 53, 62, 74  
Lohnes, P.R. 165  
Lückert, H.R. 10, 30f., 34f., 48, 120  
Lüer, G. 185

Mansholt, E. 207  
Marsolek, T. 80, 156  
Martin, E. 28, 144, 150f., 197, 238ff.,  
244  
Massialas, B.E. 23, 218  
McCarthy, D. 41f., 52  
McClelland, D.C. 35  
Meili, R. 15ff., 24f., 28f., 64, 247  
Meumann, E. 56  
Michel, L. 58f., 61, 64, 66, 74f., 79  
Mierke, K. 11f., 28, 113, 143, 161, 247  
Mittenecker, E. 144f., 155, 204  
Mühle, G. 10f., 23, 218

Neumann, H. 45, 125, 197  
Newman, H.H. 46  
Newman, J.R. 46, 195  
Nickel, H. 33f., 53  
Niepold, W. 41  
Nollau, W. 169

Oevermann, U. 34, 40ff., 48ff.

Oléron, P. 45

Orlik, P. 78

Osgood, C.E. 52

Pauli, R. 28, 65, 113, 129, 136, 143f.,  
152f., 242ff.

Pawlik, K. 14, 67, 121f., 131, 135,  
144f., 153

Peisert, H. 195

Piaget, J. 10, 16, 45

Pongratz, L. 31

Priester, H.J. 80, 114ff., 123, 211, 214

Raatz, U. 82, 128

Rapaport, D. 116

Rath, W. 213f.

Rasch, G. 55

Raven, J.C. 25, 65, 132ff., 141, 199ff.

Reinartz, A. 70, 187f.

Révész, G. 10ff.

Rexhausen, J. 207, 215

Rich, C.C. 135

Riegel, R.M. 118

Ritter, H. 33

Rorschach, H. 58, 75, 97, 145

Rosen, B.C. 36

Rösler, H.D. 80

Roth, E. 11, 14

Roth, H. 10, 12

Rousey, C.L. 45

Sader, M. 61

Samstag, K. 80

Sapir, E. 43ff., 49

Sapon, S.M. 147

Sarason, S.B. 61

Sargent, E.W. 25

Sayler, W. 155

Scheler, M. 9f.

Schell, C. 23, 218

Schenk-Danzinger, L. 10

Schirmer, B. 146f., 215f.

Schmalohr, E. 49, 186f.

Schmidt, J. 123, 207

Schollmeyer, B. 148, 197, 203f.

Schorre, G. 21

Schraml, W. 89

Schubenz, S. 95, 108, 205

Schulte, K. 48f.

Schultze, W. 164

Schwarz, E. 128

Schweingruber, W. 215

Schwinger, L. 135

Séguin, E. 56

Seifart, O. 70, 187

Seifert, K.H. 134, 197, 199

Seyfried, H. 144, 239ff.

Shurrager, H.S. 106, 123ff.

Shurrager, P.S. 106, 123ff.

Simon, T. 56, 79, 81, 207

Simmat, W.E. 155

Skowronek, H. 23

Snijders, J.T. 25, 125ff., 198ff.

Snijders-Oomen, N. 25, 125ff., 198ff.

Solomon, P. 31

Sommer, G. 94

Spearman, C. 14f., 17f., 33, 64, 114, 123,  
129, 131, 133, 247

Spitzmüller, R. 34, 36

Spitznagel, A. 61

Spreen, O. 140

Stern, W. 9ff., 56, 58, 64, 79, 81, 207

Strauß, M. 141f.

Strehle, W. 207, 215

Süllwold, F. 11, 70

Templin, M.C. 41f., 45

Tent, L. 18, 69, 81, 129, 137f., 158,  
164, 169

Terman, L.M. 78, 215

Teschner, W.P. 53, 55, 63

Tewes, U. 61

Thomae, H. 10, 86, 95

Thomas, L. 41f.

Thorndike, E.L. 14, 17

Thurner, F. 61

Thurstone, L.L. 14, 16ff., 24ff., 29, 32f.,  
113, 135f., 139, 149, 247

Tillman, M.H. 213

Toman, W. 144, 204

Trembl, F. 246

Überla, K. 14, 165

Ulich, E. 153, 243

Ulmann, G. 23, 218

Undeutsch, U. 164

Vernon, P.E. 14f., 29

Vukovich, A.F. 62

Wappmann, H. 215

Weber, E. 165

Wechsler, D. 9f., 28, 32f., 65, 69f., 80,  
100, 114ff., 118ff., 124f., 132, 141,  
160, 189, 196, 206ff., 245

Weinert, F. 34

Weingardt, E. 32

Weiß, R. 130ff.

Wendeler, J. 63, 82

Wendt, D. 14, 67

Wenzl, A. 10, 12f., 15, 247

Wertheimer, M. 10, 23

Wewetzer, K.H. 14, 29, 80, 119f. 131,  
134f., 213

Whorf, B.L. 43ff., 49  
Wilde, K. 64  
Williams, M. 100  
Wittmann, J. 48  
Witzlack, G. 80  
Wolf, R.M. 35  
Wooden, H.Z. 45

Wundt, W. 56ff.  
Zeckel, A. 197  
Zevin, J. 23, 218  
Ziehen, T. 45, 197, 207  
Zimmermann, K.W. 123, 185ff., 189,  
196, 220

Abbauindex 122

Abstraktionsfähigkeit (HAWIE/HAWIK)  
115ff., (SON) 126, (AzN) 129,  
(PMT) 133, (LPS/PSB) 136ff.

Abweichungsindex (PIT) 145

Accuracy (PMT) 133, (LPS) 137

Achievement 82

Adaptative Flexibilität 23

Äquivalentnormen 79

Äquivalenzkoeffizient s. Reliabilität

Aktivierung (der Begabung bzw. Begabungsreserven) 47ff., 161ff.

Allgemeine Begabung s. Begabung

Allgemeines Wissen (AW) 114f.

Altersnormen s. Testnormen

American Psychological Association  
(APA) 73

Anagrammtest (LPS) 135f.

Analogien (SON) 125, (AzN) 129,  
(CFT) 130f., (PMT) 133, (LPS) 135

Analogieschlußdenken 133

Analyse, statistische (Tests) 61ff.

Anamnese s. psychodiagn. Gespräch  
– (Protokoll-)Bogen 87f.

Anarchic Doctrine 14

Anforderungsmaßstäbe, schulische (test-psychologische) 162f., 166ff., 176

Angst, Ängstlichkeit 46, 60f., 81

– Prüfungs- 60, 85  
– Test- 60f.

Anlage-Umwelt-Problematik 10ff., 33f.,  
161f.

Anpassungsvermögen (sensu Meili) 15f.

Anschaulicher Zusammenhang (SON) 125f.

Anstrengungsbereitschaft 113, 137, 152f.

Antwort-(Mehrfach-)Wahl 130ff., 140,  
146ff.

Aphasiensyndrom (HAWIE) 118

Approximative Expertenabstimmung  
(Schuleignungsermittlung) 164

Aptitudes (Kreativitätsforschung) 22f.

Arbeits/haltung 28, 137, 152f., 161, 183  
– leistung 28, 137, 152f., 161f.

– probe (LPS) 28, 136f.

– qualität 137, 152f.

– sorgfalt 137, 152f., 161f.

– verhalten 28, 113, 137, 143, 150ff.,  
161f., 166ff.

– verlauf 136f., 143f., 153

Assoziation(en) 18, 20ff., 41, 115f.

– heterogene s. syntagmatische

– homogene s. paradigmatische

– paradigmatische 41

– syntagmatische 41

Assoziative Flüssigkeit (Kreativitäts-  
forschung) 21ff.

Attitudes (Kreativitätsforschung) 22

Auffassungs/geschwindigkeit (-leichtig-  
keit) 19

Aufgaben/analyse (Itemanalyse) 61ff.

– zum Nachdenken (AzN) 128ff., 167ff.,  
183f., 195f.

Aufmerksamkeit 11f., 22, 150ff., 161

Ausdauer 137, 152f.

Außenkriterium (Testvalidierung) 72ff.

Außerintellektuelle Faktoren (der Intelli-  
genzleistung) 31ff.

Ausdrucksflüssigkeit (Fluency-of-expression)  
147f.

Ausstattung, intellektuelle 12

Automatische Klassifikation (AUKL)  
98, 165ff.

Bead Arithmetic (HISAB) 124

Bedingungen der Testdurchführung

– äußere 60

– innere 60f.

Bedingungen der Intelligenz(leistung)

– soziokulturelle 30ff., 161f.

– sozioökonomische 36f.

“Bedürftigkeit” s. Sonderschul-

Begaben 13, 30ff., 35ff., 47ff., 51ff., 161f.

Begabung

– allgemeine 14f., 113ff., 124ff., 131,  
133, 136, 138, 163f., 170ff.

– Anpassungsfunktionen der 9ff., 161

– biogenetischer Anteil der 11ff., 161

– Definition(en) der 9ff., 161f.

– dynamischer Charakter der 11ff.,  
31ff., 161f.

– Hilfs- oder Stützfunktionen der 11f.,  
161 (s. auch Anpassungsfunktionen  
der Begabung)

– Sonder- 10f.

– Technische 137, 139, 149f., 170ff.,  
192

– und Intelligenz 10ff.

Begabungs/anlagen 161f.

– diagnostik 17, 81ff., 112ff., 157ff.,  
218ff., 222ff.

– differenzierung 17ff., 81ff., 112ff.,  
138, 157ff., 164ff.

– entfaltung 35ff., 47ff., 161ff.

– entwicklung 47

– ermittlung s. -diagnostik

– forschung 9ff., 29f., 31ff.

– funktionen (sensu Gottschaldt) 12f.

**Begabungs/kapazität (sensu Wenzl u. Gottschaldt)** 12

- klassifikation s. -differenzierung
- kriterien, schulrelevante 166ff.
- leistung 34, 161ff.
- merkmale, spezielle 113, 166ff.
- niveau 186f., 189, 227ff.
- potentiale (-potenzen) 132f.
- prognose 32f., 161ff. (s. auch -diagnostik)
- reserven s. Bildungsreserven
- richtung 10f.
- tests, allgemeine 114ff.
- tests, differentielle 135ff.
- tests, spezielle 140ff.
- testsystem (BTS) 148
- theorien (-modelle) 9ff.
- untersuchungen 157ff.
- vergrößerung 31ff., 161f.

**Begriffs/bildung** 44ff., 54

- schatz (-umfang) 115f., 129, 136, 139, 146f.

**Belastbarkeit** 137, 143, 150ff., 161

**Benennung (Naming factor)** 147

**Benton-Test** 140f., 202, 229f.

**Beobachtung** s. Verhaltensbeobachtung

**Beratung, informative** 95ff. (s. auch Berufs- und Bildungsberatung)

**Beratungsinteraktion** s. Approximative Expertenabstimmung (sensu Aurin)

**Berliner Englisch-Leistungstest** 148

**Berufs/beratung** 81ff., 154f.

- eignungsfindung 141ff., 149ff., 155
- erfolg Blinder 141ff.
- Interessenrichtungen 154f.
- Interessen-Test (BIT) 78, 113, 145, 154ff., 204f.

**Beurteilungs/bogen (Schüler)** 86ff.

- maßstäbe (Schuleignungsfindung) 166ff.

**Beweglichkeit** s. Plastizität

**Bewertung (Denkoperation)** 19ff., 28f.

**Bewertungsmaßstäbe der Schule** 162ff.

**Beziehungen** s. Denkprodukte

**Beziehungserfassung** 131ff., 207

**Bild. Kunst, Interessen für (Test)** 145

**Bilderergänzen (BE)** 116, 124

**Bilderordnen (BO)** 116

**Bilder-Testserie (BT 1-2, BT 2-3)** 220

**Bildungs/beratung** 81ff., 86ff., 95ff.

- chancen 30f., 35ff., 47ff.
- einflüsse 35ff., 161f.
- empfehlung 164ff.
- erfolg 161ff., 220
- förderung (vorschulische) 33f., 47ff., 82f.
- forschung 35ff., 82ff., 111ff., 218ff.
- interessen 161f., 169f., 219
- maßnahmen (Begabungsförderung) 35ff., 47ff., 51ff., 162
- reformen und Begabung 163

**Bildungs/reserven** 12, 47ff., 88

- standard 135f.
- ziele 161f.

**Binet-Skalen (Binetarium)** 32f., 119ff., 125f., 134, 207, 213

**Blinde** 112, 120, 206ff.

- lernbehinderte 213f.

**Blinden/psychologie** 207

- diagnostik 123ff., 141ff., 146f., 207ff.

**Block Design (HISAB)** 123

**Bourdon-Test** 113, 150f., 238

**California Test of Mental Maturity** 120

**Ceiling-Effekt** 116

**Centil-Werte (C)** 76ff., 136ff., 245

**Charkow-Test** 25

**Classes (Klassen)** s. Denkprodukte

**Clear Thinking (PMT)** 133

**Closure** 19, 137, 139

**Code** s. Sprachkode

**Cognition (Denkoperation)** 19f., 28  
(s. auch Kognition)

**Conditioned factors (sensu Cattell)** 29

**Convergent Production** 19f.

**Creativity** s. Kreativität

**Culture Fair or Free Intelligence Tests (CFT)** 130ff., 195f.

**Cut-off-Score (Cutting Score)** 164, 185

**Datenerhebung (-erfassung)** 83ff., 162ff.

**Debität** s. Lernbehinderung

**Deduktion** 18, 133

**Denken (Denkfähigkeit)**

- abstrakt-logisches (schlußfolgerndes) 11ff., 18ff., 25ff., 115f., 126, 129, 133, 136, 139, 142, 146, 149

- anschauungsgebundenes 25f.

- beziehungsstiftendes 131ff.

- deduktives s. Deduktion

- formallogisches 18, 27f. (s. auch abstrakt-logisches)

- induktives s. Induktion

- produktives 19ff.

- sprachgebundenes 18, 27ff., 135ff., 146ff.

- technisches (technisch-konstruktives) 19, 137, 149f.

- zahlengebundenes 18, 27f., 115ff.

**Denkinhalte (sensu Guilford)** 20f.

- behaviorale (Verhaltens-) s. soziale

- bildliche 20f.

- figurale 20f.

- semantische 20f.

- soziale 20f.

- symbolische 20f.

**Denkoperationen (sensu Guilford)** 19ff.

- bewertende 19f.

- Denkoperationen, kognitive 19f.
  - produktive 19f.
- Denkprodukte (sensu Guilford) 20 ff.
  - Beziehungen (Relations) 20f.
  - Einheiten (Units) 20f.
  - Implikationen (Implications) 20f.
  - Klassen (Classes) 20f.
  - Systeme (Systems) 20f.
  - Transformationen (Transformations) 20f.
- Denk/prozesse 16, 18ff.
  - divergierende s. Divergent Production
  - konvergierende s. Convergent Production
  - psychologie 11, 19ff.
  - verhalten, intuitives 12f., 21ff.
- Depression (Test) 144
- Deprivation
  - sensorische 31, 53, 80f.
  - soziale (sozio-kulturelle) 31, 33, 40, 46, 53, 81
  - sprachliche 43, 46
- Deterioration Index s. Abbauindex
- Determinanten, sozio-kulturelle 35ff., 163f.
- Determinismushypothese, sprachliche 44f.
- Dezibel (db oder dB) 196
- DFT s. Figural adaptative Flexibilität
- Diagnostik, pädagogische 81ff.
- Diagnostische Valenz s. Validität
- Differenzen
  - interindividuelle 56ff., 159ff.
  - intraindividuelle 159ff.
  - kritische 157, 186ff.
  - matrizes 157ff., 194f.
  - Testleistungs- 170ff.
- Digit Symbol (HISAB) 123f.
- Dimensionen, psychologische (der Intelligenz) 14ff., 114ff., 173
- Diskriminanzanalyse, multiple 165
- Divergent Production (Divergente Produktion) 19ff., 25ff.
- Diversifikationsquotient 40, 53
- DMC s. Semantische spontane Flexibilität
- DMI s. Elaboration
- DMR s. Assoziative Flüssigkeit
- DMT s. Originalität
- DMU s. Gedankenflüssigkeit
- DP s. Divergent Production
- DP-Faktoren 21
- DP-Tests 21
- DSS s. Divergente Produktion
- DSU s. Word Fluency
- Dynamik (Intelligenz) 11ff., 31ff.
- Eichung s. Testeichung
- Eignung 10f., 161ff.
  - Berufs- 11
  - Gymnasial- 166ff.
  - Hauptschul- 166ff.
- Eignung, Realschul- 166ff.
  - Schul- 10f., 17, 21, 166ff.
  - Studien- 21ff.
- Eignungs/ermittlung 71f., 161ff., (Fallbeispiele) 176ff.
  - kriterien, quantitative 162ff.
  - profile 167ff.
- Einfallsreichtum 25
- Einheiten (Denkprodukte) 20f.
- Einseitige Fragestellung 70f.
- Einstellung, soziale (Test) 144
- Elaboration (Kreativitätsforschung) 23
  - syntaktische 41
- Endothymen Grund und Begabung 12f.
- Englisch-Leistungs-Test 148
- Englische Schule (Faktorenmodell) 14f., 29
- Entscheidung, relative Wahl- (Test) s. ipsative Werte
- Entwicklung, geistige 31ff.
- Entwicklungs/alter 79, 126f.
  - stand, sachstruktureller 48f.
  - test s. Test
- Erbe-Umwelt-Bedingungen (Begabung) 33ff.
- Erfahrung und Sehen (Wahrnehmung) 208
- Erfassungsdimensionen (Tests) (HAWIE) 114ff., (HAWIK) 119, (HISAB) 124, (SON) 125f., (AzN) 129, (CFT) 131, (PMT) 133, (LPS) 136f., (PSB) 139, (Benton) 140, (VISAB) 142, (KVT) 143, (PIT) 145, (WST) 146, (FTU) 147, (MTVT) 149, (Bourdon) 150, (KLT) 152, (Pauli) 153, (BIT) 154
- Ergänzung (SON) 125, (MTVT) 149
  - Lückentests 20, 147
- Erkenntnis oder Erfassung (Denkoperation) 19f.
- Erlebnisbeobachtung 84
- Erziehung, kompensatorische 46ff.
  - Interessen für (Test) s. Einstellung, soziale
- Erziehungs/einflüsse, familiäre 35ff., 47ff., 205
  - schwierige 112
  - stile s. Kindererziehung
- Evaluation (Bewertung) 19f., 28f.
- Experiment
  - Forschungs- 58
  - Prüf- oder diagnostisches 58
- Expertenabstimmung, approximative 164
- Exploration s. psychodiagn. Gespräch
- Expression, Fluency-of 147
- Extraversion (Test) 144
- Fähigkeit(en)
  - differentielle 17ff., 136ff.
  - kognitive s. Denken/Intelligenz
  - kombinatorische 125ff., 131ff.

- Fähigkeit(en), Primär- 17ff., 202
- rechnerische 18, 27f., 115, 119, 124, 129, 137, 139, 153
- sprachlogische 129
- technisch-konstruktive 19, 137, 149f., 217
- verbale 18, 27ff., 38ff., 114ff., 129ff., 135ff., 146ff.
- Fähigkeits/richtungen 17
- tests s. Tests
- Faktoren/analyse 14ff., 17ff., 24ff., 74
- außerintellektuelle 31
- modelle (-theorien) 14ff.
- struktur (HAWIE/HAWIK) 114ff., (LPS/PSB) 136ff., (FTU) 147f., (SON) 125f., 197ff.
- Fehlanpassung, neurotische (Test) 144
- Fehler/bemerken (LPS) 137
- vom Typ Alpha s. Risiko erster Art
- vom Typ Beta s. Risiko zweiter Art
- Figural adaptative Flexibilität 21
- Figuren/Durchstreich-Test (Bourdon) 28, 150f., 238
- Legen (FL) 116, 124
- Figure Reasoning Test 220
- Findewertigkeit s. Validität
- Findigkeit, theoretisch-technische 137
- Flächentransformation 77f.
- Flexibilität 21ff.
- adaptative 21, 23
- Flexibility of Closure 19, 137
- Fluency (Flüssigkeit) 15f., 18, 21ff., 25ff., 136, 147
- Form (SON) 125
- bretter 133f.
- legetest 150
- sehen und Denken 208
- Frage/bogen 83, 86ff.
- stellung (einseitige vs. zweiseitige) 70f.
- Frankfurter Wortschatztest (WST) 52f., 65, 78, 103, 113, 129, 146f., 202f., 216, 234ff.
- Fremdsprachen/eignung 147ff., 182ff.
- Eignungstest (FTU) 71, 78, 113, 147ff., 182ff., 202ff.
- Gaußsche Kurve s. Normalkurve
- Gedächtnis 11f., 18ff., 27ff., 115, 118, 123, 148, 161, 215, 219
- assoziatives 147
- faktor, undifferenzierter (HAWIE) 118
- für Karten (SON) 125
- unmittelbares 126
- Gedankenflüssigkeit 21
- Gehörlose 43, 45f., 51ff., 112, 126ff., 132ff., 144, 146f., 151f., 155, 196ff., 220, 222ff.
- Gehörlosen/psychologie 11
- schule 196f., 202f.
- Gemeinsamkeitenfinden (GF) 115
- Generalfaktor 14ff.
- Genetisches Potential 34
- Geschlechtsunterschiede (Kreativität) 21
- Gestalt/ergänzung (-entwicklung) 135f.
- psychologie (Denken) 15, 20
- wahrnehmung 135, 137
- Globalisation 15f., 19
- Grammatical Sensitivity 148
- Grammatikunterricht 51
- Grenzwertmethode 164f., 185f.
- Grobnormen 80, 230, 238f., 242ff.
- Grund/gesamtheit 67
- schulunterricht (Begaben) 50f.
- wert von Testbefunden 97f.
- Gruppen/faktoren 15ff.
- test (-verfahren) 128ff., 146ff.
- Gültigkeit s. Validität
- Gutachten, pädagogisches 94ff., (Fallbeispiele) 98ff.
- beratendes 95
- Persönlichkeits- 95
- Problem- 95
- technik 96ff.
- übermittlung 109f.
- Gymnasialeignung 166ff.
- Halbierungsmethode 65
- Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE) 25, 28, 65, 78, 106, 114ff., 124f., 205f.
- Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK) 25, 28, 32, 65, 76, 78, 100ff., 119ff., 128, 134, 141, 143, 185ff., 192, 196, 200ff., 231ff.
- Handlungsfaktor (HAWIE/HAWIK) 118 f, 185
- Handwerk, Interessen für (Test) 145, 154
- Haptic Intelligence Scale for Adult Blind (HISAB) 106, 123ff.
- Heilpädagogik 162
- Hirnorganische Störungen 140f.
- Hör-(Sprach-)geschädigte (-behinderte) 43, 45ff., 51ff., 112, 120, 125ff., 132ff., 140f., 151, 153, 196ff.
- Hör-(Sprach-)geschädigtendiagnostik 112f., 120, 125ff., 132ff., 140, 144, 146ff., 196ff., 219f., 222ff.
- Illinois Test of Psycholinguistic Abilities (ITPA) 52
- Implications (Denkprodukte) 20f.
- Induktion 18, 133
- Insight-Test 25
- Instruktionsverständnis (IV) 129



## Integration (Kreativitätsforschung) 22

### Intelligenz/abbau 32, 114ff., 122

- allgemeine 14f., 114, 123, 131ff., 140f., 202
- alter 79, 126
- Anpassungsfunktionen der 161
- bedingte Faktoren der (sensu Meili) 16
- definitionen 9ff.
- diagnostik 31ff., 112ff. (s. auch Begabungsdiagnostik)
- disposition 29
- und Denken 11
- dynamisches Modell der 15
- entwicklung s. Ontogenese der
- faktoren 13ff., 17ff., 218f.
- Faktorentheorien (Modelle) der 14ff.
- forschung, operationale 29f.
- forschung, traditionelle 11
- funktionen 11ff., 20ff., 40, 114ff.
- Generalfaktor (g-factor) 14f., 114ff.
- Grundfaktoren der 16ff.
- Grund- 11f., 131
- Gruppenfaktoren der 17ff.
- Handlungs- 114ff., 141ff., 149f., 208ff., 212f.
- Hauptdimensionen (Faktoren) der 14ff.
- hierarchisches Gefüge der 11f., 14f., 29
- Hilfs- oder Stützfunktionen der 11f., 28, 113, 144, 150ff., 163f., 218f.
- Instrumentalcharakter der 10
- kapazität 115f.
- Kern- 11f., 28, 51, 113, 218f.
- Konzepte, multifaktorielle der 17ff.
- Kurve (Ontogenese) 32f.
- kulturunabhängige 130ff.
- messung, nonverbale 45, 125ff., 130ff., 141ff., 197ff.
- niveau 32, 96
- Ontogenese der 31ff., 42f.
- praktische (technische) 27, 149f.
- Primärfähigkeiten der 17ff.
- Punkt-Test (IPT) für Blinde 207
- Quotient (Abweichungs- oder Wechsler-IQ) 78f., 118f., 245
- Quotient (Sternscher IQ) 79
- Sekundär- 113, 143, 152
- soziale 20
- Struktur 12f., 15ff., (Modell) 19ff., 135ff.
- Struktur-Test (IST) 25, 131, 152
- Stützfunktionen der 11f., 28, 51, 113, 155f., 161, 164, 170, 218f.
- temperament 12
- testanwendung 81ff.
- test für normalsichtige und sehgeschädigte Kinder und Jugendliche (INS) 100
- tests 22, 94f., 112ff., 218f. (s. auch Begabungstests)
- tests, allgemeine 113ff.

### Intelligenz/tests, differentielle 113, 135ff.

- tests, kulturunabhängige 130ff.
- tests, nonverbale 45, 125ff., 130ff., 141ff.
- tests, spezielle 140ff.
- theorien, faktorenanalytische 14ff.
- theorien, phänomenologische 10ff., 15, 29f.
- Verbal- 114f., 202ff., 209ff. (s. auch verbale Fähigkeiten)
- Würfelmodell (morphologisches Modell) 20, 24
- Interferenzneigung (Test) 131
- Interim-Hayes-Binet-Tests for the Blind (IHB) 207
- Interessen 12, 144ff., 154ff., 169f., 204f.
  - als Hilfsfunktion der Intelligenz 11f.
  - bildungs- und berufsrelevante 145f., 154ff., 163f.
  - tests 144ff., 154ff.
  - variablen s. -tests
- Interkorrelation (Tests) 158ff.
- Intervallskala 77f.
- Introversion (Test) 143
- Invarianz des IQ 31ff.
- IQ s. Intelligenzquotient
- Ipsative Items 144f., 154f.
  - Werte 154f.
- Irrtumswahrscheinlichkeiten (Testinterpretation) 68ff.
- Item/analyse s. Aufgabenanalyse
  - auswahl (Testreliabilität) 65f.

### Kaufmännische Interessen (Test) 145, 154

### Kinder-Angst-Test (KAT) 61

### Kinder/erziehung, Formen der 36ff., 46

### – gartenpädagogik 49

### Klassen (Denkprodukte) 20f.

### Klassifikation, automatische (AUKL)

98, 165ff.

### Klassifikations/modell 164f.

### – probleme 163ff., 185ff.

### – skala 77

### – verfahren (Schuleignungsermittlung) 165ff.

### K : m 15

### Knox Würfel (SON) 125

### Kode s. Sprachkode

### Kognition (Cognition) 27ff., 42f., 45ff.

### Kognitive Denkfunktionen (Faktoren) 19f., 28f.

### Kollektiv s. Grundgesamtheit

### Kombination (SON) 125

### – Fähigkeit zur 149f.

### Kompensationshypothese (sensu Klauer) 211ff., 216

### Komplexität 15f., 26ff.

### Konfidenzintervall 67ff., 157ff.

**Konsequenz-Test** 25  
**Konsistenz (innere) s. Reliabilität**  
 – **koeffizient** 157  
 – **Inter-Item- (-Analyse)** 65f., 163  
**Konstanz, intraindividuelle** 31ff., 161f.  
 – **IQ-** 31ff.  
**Konstrukt(validität) s. Validität**  
**Konvergente Produktion** 19f., 28  
**Konvergenzhypothese** 13  
**Konzentration** 11, 18, 25ff., 113, 115f., 137, 143f., 150ff., 161f.  
**Konzentrations-Leistungs-Test (KLT)** 113, 151f.  
**Konzentrations-Verlaufs-Test (KVT)** 28, 113, 143f., 239 ff.  
**Konzeptbildung** 45, 54, 214  
**Körperbehinderte** 112  
**Korrelation (Koeffizient)** 32  
 – **(punkt)biseriale** 62, 129, 138  
 – **triseriale** 138  
**Kraepelinsche Arbeitsprobe** 152f.  
**Kreativität** 21ff., 75, 218  
**Kreativitäts/forschung** 21ff.  
 – **tests** 18ff., 75, 218  
**Kretschmer-Höhn-Test** 129  
**Kritische Periode s. sensible Phase**  
**Kuder-Richardson-Methode (Konsistenz-analyse)** 129, 142, 154  
**Kurzzeitgedächtnis (Lb-Symptom)** 185  
  
**Labil, vegetativ (Test)** 144  
**Längsschnittuntersuchungen (Longitudinalstudien) der Intelligenzentwicklung** 31ff.  
**Landleben, Interesse für (Test)** 145, 154  
**Learning Ability, Inductive Language (FTU)** 148  
**Lebenspraktisch Bildungsfähige** 112  
**Legasthenie, Legastheniker** 82, 106ff., 112, 137, 151, 205f.  
**Leistungs/beurteilung (-bewertung), schulische** 163f., 202, 220  
 – **druck** 36  
 – **fähigkeit, allgemeine (intellektuelle)** 131ff., 152f.  
 – **fähigkeit, sprachliche** 114ff., 128ff., 135ff., 146ff.  
 – **funktionen, intellektuelle** 161ff.  
 – **menge** 151ff. (s. auch Arbeitsverhalten)  
 – **motivation** 11f., 18, 34ff., 113, 137, 151ff., 161, 218f.  
 – **motivation, Aufbau der** 35f., 48ff., 161f.  
 – **potential** 137  
 – **Prüf-System (LPS)** 25, 28, 65, 69, 78, 113, 135ff., 150, 157ff., 166ff.  
 – **qualität, Verlauf der** 143  
 – **schwäche, hirnorganisch bedingte** 116ff., 135, 140f.

**Leistungsverhalten** 31, 161f., 166ff.  
 (s. auch Arbeitsverhalten)  
**Leiter International Performance Scale** 220  
**Lern/behinderung, Lernbehinderte** 90ff., 98ff., 106ff., 112, 185ff., 213f.  
 – **behindertendiagnostik** 112, 118, 120, 135, 185ff., 213f., (Fallbeispiele) 90ff., 98ff., 106ff.  
 – **fähigkeit** 115f., 129, 161f., 185f.  
 – **fortschritte (Erfassung)** 51ff.  
**Lernen, schulisches (Test)** 137  
**Lesebuchreform** 48  
**Lesenlernen, Fähigkeit zum** 48, 50f.  
**Lese-Rechtschreib-Schwäche (LRS) s. Legasthenie**  
**Linguistische Relativität, Hypothese der** 44  
**Literarische Interessen (Test) s. sprachliche Interessen**  
**Logische Beziehungen** 20  
**Logisches Schließen s. Denken, formallog.**  
**Lückentest** 20, 56  
**Lügentest** 60f., 146  
  
**Major group factors** 14f.  
**Manisch (Test)** 144  
**Manual ability** 15  
**Mathematikbegabung** 16, 18, 27f., 115, 124, 129, 137  
**Matrizen (Tests)** 131ff.  
**Mechanical Aptitude Battery** 149  
 – **information** 15  
**Mechanisch-Technischer Verständnis-Test (MTVT)** 113, 149f.  
**Mehrfachbehinderung** 206, 216f.  
**Mehrfachwahl-Antwortform (Multiple Choice)** 130ff., 140, 146ff.  
**Memory** 11ff., 18ff., 26ff., 115f., 125f., 161f., 185  
 – **associative** 147  
 – **Rote, for Foreign Language Materials** 148  
**Merkfähigkeit s. Memory**  
**Merkmalsstabilität, intraindividuelle** 31ff., 161ff.  
**Merkwelt (sensu Uexküll)** 12  
**Messen** 77ff.  
**Meßfehler s. Standardmeßfehler**  
 – **des Median** 70  
 – **für Prozentränge (%-Zahlen)** 69  
 – **für Quartile** 70  
 – **für Standardwerte** 69  
 – **Konzept (Testinterpretation)** 157ff.  
 – **Theorie** 67f.  
**Meßgütekriterien (Testgütekriterien)** 63ff., 162  
**Mill-Hill-Vocabulary-Test** 132  
**Milieubedingungen, sozio-kulturelle (der Begabung)** 30ff., 161f.

Milieu/einflüsse 33ff., 38ff., 51ff., 114ff.,  
 130f., 136f., 146f., 195f., 205, 216f.  
 – variablen s. -bedingungen (sozio-kultu-  
 relle Determinanten) der Begabung  
 Minderungskorrektur (Validität) 74  
 Minor group factors 14f.  
 Modern Language Aptitude Test 147  
 Montessori-Pädagogik 49  
 Mosaik (SON) 125  
 Mosaik-Test (MT) 25, 116ff., 123f., 134  
 Motivation und Testleistung 60f., 163f.  
 (s. auch Leistungsmotivation)  
 Motivationsbarrieren 60  
 Motive als Hilfsfunktion der Intelligenz  
 11f., 26ff.  
 Multiple Choice s. Mehrfachwahl-Antwort-  
 form (Tests)  
 Musikalische Interessen (Test) 145  
  
 Naming factor 147  
 Neurotisch (Test) 144  
 Noetischer Oberbau (sensu Lersch) 13  
 Non-aptitude-Traits (Kreativitätsforschung)  
 22f.  
 Normen s. Testnormen  
 Normalisierung 77f.  
 Normalkurve 66  
 Normative Werte 154  
 Number 16ff., 26ff., 115, 124, 129, 137,  
 192 (s. auch Numerical ability)  
 Numerical ability 15, 214  
  
 Oberschulsyndrom (HAWIE) 118  
 Object Assembly (HISAB) 124  
 – Completion (HISAB) 124  
 Objektivität 59ff., 63, 72f.,  
 (Tests) 114ff., 163  
 Observational Techniques 84ff.  
 Ökonomie (Test) 9f., 132, 135, 147, 218f.  
 Oligarchic Doctrine 14  
 Operationismus 9f., 13ff., 29ff.  
 Ordinalskala 77f.  
 Ordnen von Begriffen (Convergent  
 Thinking) 20  
 Organikersyndrom (HAWIE) 118  
 Originalität 21ff.  
 Osgood-Modell sprachl. Kommunikations-  
 systeme 52  
 Overachievement 82, 85, 163  
 Overlapping 165  
  
 Papier-Bleistift-Verfahren (Paper and  
 Pencil Tests) 130ff.  
 Paralleltest-Methode (Reliabilität) 65

Parameter 66f.  
 Paranoid (Test) 144  
 Pattern Board (HISAB) 124  
 Pauli-Test 28, 65, 113, 129, 152ff., 242ff.  
 Perceptual Speed 19, 137, 139  
 Performance 29, 33, 114ff., 123ff.,  
 130ff., 141ff.  
 Peristase (Umwelt) 34ff. (s. auch Milieu-  
 bedingungen bzw. -einflüsse)  
 Persönliche Gleichung 57  
 Persönlichkeits/faktoren (Test) 144  
 – Fragebogen 144ff.  
 – und Interessen-Test (PIT) 78f., 113,  
 144ff., 154ff., 169f., 179ff., 204f.  
 – variablen 22, 34, 46 (s. auch PIT  
 u. BIT)  
 Phantasie 21ff., 26ff., 161  
 Phonetic Coding 147  
 Placement 81  
 Plastizität 15f., 19  
 Population s. Grundgesamtheit  
 Potential, genetisches (Begabung) 34  
 Powertest s. Test  
 Prädiktorvariablen 162ff., 169f., 182ff.  
 Primärfähigkeiten (Intelligenz) 17ff.,  
 32, 136f., 139  
 Primary Mental Abilities (PMA) 135  
 (s. auch Primärfähigkeiten)  
 Prinzipien s. Abstraktionsfähigkeit  
 Problemlösung 18ff.  
 Produktion 19f.  
 Produktives Denken 19ff., 23, 25ff.  
 Profil/analyse (-auswertung) 119ff.,  
 125ff., 129f., 138ff., 157ff., 167ff., 187f.  
 – differenzen 157ff.  
 – Gruppen- 160  
 – Individual- 160  
 – interpretation s. -analyse  
 – reliabilität 137f., 158ff.  
 – zeichnung 136, 170ff.  
 Prognosegültigkeit s. Validität  
 Progression (der Aufgabenschwierigkeit  
 im PMT) 132  
 Progressive Matrices s. Progressiver  
 Matrizen-Test  
 – – Advanced 132  
 – – Coloured 132  
 – – Standard 132ff.  
 Progressiver Matrizen-Test (PMT) 25,  
 61, 65, 69, 78, 128, 132ff., 199ff., 222ff.  
 Prozenstränge (PR) 69ff., 76ff., 245  
 Prozeßfaktoren (Intelligenz) 16f.  
 Prüf-System für Schul- und Bildungs-  
 beratung (PSB) 65, 78, 131, 139f.,  
 148, 160, 167ff., 189ff.  
 Psycho/linguistische Fähigkeiten 52  
 – linguistische Forschung 38ff.

Psychodiagnostisches Gespräch 83f.,  
89ff., 152

Rangskala s. Prozenträge

Ratefähigkeit (Test) 137, 139

Rationalskala 77

Raumvorstellung (Räumlichkeitsfaktor)

16, 19, 116, 123f., 133, 137, 139

Reagieren, richtiges (Test) 137, 139,  
152f.

Reasoning (Denken) 18ff., 25ff., 115ff.,  
129, 131ff., 136f., 139, 142, 149, 170,  
192, 202

– Arithmetic 26ff., 115, 124, 213 (s. auch  
Rechenfähigkeit)

– Verbal 16, 18, 20ff., 25ff., 114ff.,  
119ff., 129f., 135ff., 139f., 146ff.,  
166ff., 202ff., 208ff., 216

Rechen/fähigkeit (-fertigkeit) 16, 18, 27f.  
115, 124, 129, 137

– operationen 18

Rechnen (RE) 129

Rechnerisches Denken (RD) 115

Rechtschreibkenntnisse 136f.

Redefinition 23

Regelerkennen 18, 133

Regressionsanalyse 165

Rehabilitation, berufliche (Blinder) 141ff.

Reihenfortsetzen 129, 135f., 139

Relations (Denkprodukte) 20f.

Reliabilität 63ff., 72f., (Tests) 114ff.,  
157ff., 164, 218

– Paralleltest- 65

– Profil- 157ff.

– Retest- 65, 157

– Split-half- 65

Reliabilitätskoeffizient 64f.

Repräsentanz (Testitems) 61f., 83f.,  
163f., 218f.

Resthörige (Hörrestige) 51ff., 196ff.,  
222ff.

Retardierung 39ff., (Blinder) 207f.

Retest s. Reliabilität

Reversalshift-Reaktionen (Gehörloser) 45

Reversibilität 16

Revised Visual Retention Test s. Benton-  
Test

Rigidität 22

Risiko erster Art 71

– zweiter Art 71

Sachverhaltserkennen (Blinder) 207

Sapir-Whorf-Hypothese 43ff.

Satzergänzung (SE) 129

Schizoid (Test) 144

Schöpferisches Denken 21ff., 25ff.

Schreibenlernen, Fähigkeit zum 50f.  
Schülerbeurteilung 81ff., 94ff., 157ff.,  
161ff., 182ff., 185ff., 196ff., 205f.,  
207ff., 222ff.

Schuleignung (Definition) 10ff., 161ff.

– diagnostische Erfassung der 161ff.

– Dynamik der 161f.

– Gymnasial- 162f., 166ff., 173ff.,  
213f., (Fallbeispiele) 102ff., 176ff.

– Haupt- 162f., 166ff., 171, 175ff.,  
(Fallbeispiele) 176ff.

– Konstituierung der 20ff., 161f.

– Mittel- 162f., 166ff., 172, 177,  
(Fallbeispiele) 176ff.

– Real- s. Mittel-

– Sonder- (Lernbehinderung) 162, 185ff.,  
213f., (Fallbeispiele) 98ff., 106ff.

Schuleignungs/differenzierung 17, 74f.,  
81f., 112, 118, 121f., 129ff., 136ff.,  
145ff., 154f., 157ff., 161ff., 166ff.,  
(Fallbeispiele) 176ff., 185ff., 218f.

– erfassung (-ermittlung) s. -differen-  
zierung

– kriterien 157, 162ff., 166ff., 171ff.

– prognose (-prognostik) s. -differen-  
zierung

– urteile 163ff., 177ff.

Schul/laufbahnberatung 81f., 86ff.,  
95ff., (Fallbeispiele) 176ff.

– leistung(en), Beurteilung von 47ff.,  
53, 129f., 135ff., 161ff., 169f., 220

– leistungsanalyse 81f., 112

– leistungsbeurteilung s. -leistung(en)

– leistungsfaktor 185f., 213f.

– noten, Skalenqualität der 78, 245

– reife 34, 81

– wesen, allgemeinbildendes 112

– wesen, Sonder- 112

Schwachsinnige 81, 116 (s. auch  
Lernbehinderte)

Schwachsinnigensyndrom (HAWIE) 118

Schwerhörige 43, 51ff., 112, 134f.,

146f., 196ff., 222ff.

Schwierigkeit (Testaufgaben) 61f.

Schwierigkeitsindex 61f.

Sehbehinderte 112, 120, 146f., 208ff.,  
213ff. (s. auch Sehgeschädigte)

Sehgeschädigte 88, 90ff., 98ff., 120,  
147, 207ff.

Sehgeschädigtendiagnostik 88, (Fallbei-  
spiele) 90ff., 98ff., 112, 114ff., 123ff.,  
135, 141ff., 146f., 207ff.

Selbst/beurteilung s. -kritik

– kritik (Test) 144f.

Selection (Selektion)

– Kriterien der 163ff.

Semantische spontane Flexibilität 21

Sensible Periode (– Phase) 34

- Skala, Abweichungs-IQ- 78f., 245
  - Centil- oder C- 78, 245
  - Intervall- 77ff.
  - IQ- 77ff., 245
  - Klassifikations- 77
  - Nominal- s. Klassifikations-
  - Normen- s. Testnormen
  - Ordinal- 77ff., 245
  - Rang- s. Ordinal-
  - Rational- 77
  - T- 78, 245
  - Verhältnis- s. Rational-
  - Wertpunkt- oder WP- 78, 245
  - Zensuren- 245
- Snijders-Oomen Nichtverbale Intelligenztestreihe (SON) 125ff., 197ff.
- Sonderpädagogische Diagnostik 185ff.
  - Indikation (Lernbehinderung) 185
- Sonderschul-“Bedürftigkeit” 162, 185ff.
- Sorgfaltsstreben (Test) 137, 152f.
- Sortieren (SON) 125
- Soziale Einfühlung s. Intelligenz, soziale
  - Interessen (Test) 145, 154
- Sozialisation 31 (s. auch Kindererziehung)
- Sozio-kulturelle Determinanten der Begabung s. Milieueinflüsse
- Space 16, 19, 25
- Spatial Ability 15
- Spearman-Brown-Formel (s. Halbierungsmethode) 117, 129, 148
- Special factors, specific factors (s-Faktoren) 14f.
- Speed of Closure 19
- Split-half-Reliabilität s. Halbierungsmethode
- Spontane Flexibilität 22f.
- Spontaneität (Kreativitätsforschung) 22
- Sprach/beeinträchtigung 196f.
  - beherrschung (HAWIE) 118
  - behinderung, -behinderte 112, 120, 125ff., 132ff., 196f., 220
  - entwicklung 35f., 38ff.
  - fähigkeit(en) s. Fähigkeiten, verbale
  - gebrauch, Formen des 38ff.
  - heilschulen 197
  - kode, elaborierter 38f., 50f., 53f.
  - kode, formaler s. elaborierter
  - kode, öffentlicher, s. restringierter
  - kode, restringierter 39f., 42f., 46f., 53f., 218
  - kode, schichtspezifischer 35f., 38ff., 46f., 94
  - losigkeit 196
  - psychologie 42ff.
  - schleife, Hypothese der 44f.
  - verzögerung 43, 196f.
- Sprache 35, 43ff.
  - syntaktische Struktur der 41
- Sprachliche Determinismushypothese 44
  - Interessen (Test) 145
- Stabdiagramme (Eignungskriterien) 167f.
- Stabil, vegetativ (Test) 144
- Stabilität s. Reliabilität
- Stadtleben, Interessen für (Test) 145
- Stammler 197
- Standard/abweichung 62, 66f.
  - bedingungen (Tests) 57, 59ff.
  - fehler d. arithm. Mittels 67f.
  - fehler d. Median 69f.
  - fehler d. mittl. Quartils (Median) 69f.
  - fehler von Prozenträngen (%-Zahlen) 69f.
  - isierung (Tests) 59ff., 64
  - meßfehler 61f., 66ff., (Tests) 114ff., 157
  - normen (-Äquivalente) 78
  - werte (SW) 126f.
  - wertskalen 76, 78ff., 245
- Stanford-Binet-Revision 207 (s. auch Binetarium)
- Statistiken 67
- Status, sozio-ökonomischer und Intelligenz (Sehgeschädigter) 216
- Stellenwert von Testbefunden 97
- Stichprobe (Sample) 66, 76
- Stichproben/statistik 66f.
  - verteilung 67
- Störbarkeit, funktionelle (Test) 131
- Störungen, hirnorganische (Test) 116ff., 140f.
- Stottern 197
- Strata 76
- Strukturanalyse 159ff. (s. auch Profilanalyse)
- Studieneignungsprognostik 81
- Stuttgarter Lückentest 129
- Subordinationsindex 42, 53
- Subtest 157f.
  - alter 126
- Symbolische Denkinhalte (Symbolic) 20f., 28
- Symbolverständnis 48
- Syndrom 97f.
- Syntaktische Organisation 46
- Systeme, Systems (Denkprodukte) 20f.
- Talent 10
- Taube, Taubstumme (Gehörlose) 43, 69, 125, 128, 146, 196ff.
- Taubstummenschule (Gehörlosen-) 197
- Technische (technisch-konstruktive) Befähigung 19, 137, 139, 149
  - Interessen (Test) 145, 154
- Tempo-Motivation 25f., 143, 152f.
- Tenazität 151

Test(s) 55f., 58, 113ff.

- angst 60
- angstskala 61
- auswertung 76ff., 96ff.
- batterie 157, 218
- Blinden- 114ff., 123ff., 141ff., 146f., 207ff.
- diagnostischer Prozeß 57ff.
- diagramme 177f.
- eichung 75ff.
- Entwicklungs- 133
- Fähigkeits- 135ff.
- faktorisierte 169, 219
- gütekriterien 63ff., (Tests) 114ff.
- Informelle 63
- interpretation 63, 68f., 95ff.
- kriterienbezogene s. informelle
- leistung 60f., 161
- leistungsdifferenzen 157ff.
- leiter (Anforderungen) 82f.
- normen 63, 68f., 75ff., 114ff., 222ff., 245
- für Hörgeschädigte 222ff.
- Parallel- 61f.
- Power- 65, 132, 149
- Prädiktor 164, 176
- projektive 75
- profil 114ff., 125ff., 135ff., 144f., 154f., 157ff.
- Rechen- s. Number
- skala s. Test
- Speed- 65, 150, 153
- standardisierter 55f.
- Teacher-made s. informelle Tests
- theorie 55ff., 67f., 73
- training 60
- vorform 61f.
- wiederholung 60f.
- zur Erfassung der Allgemeinbegabung 114ff.
- zur Erfassung differentieller Fähigkeiten 135ff.
- zur Erfassung spezieller Fähigkeits- und Verhaltensmerkmale 140ff.

Traits 19, 22

Transformations (Denkprodukte) 20f.

Transformation, Rohwert- 59, 76ff.

- von Testnormen und Zensuren 59, 245

Transparentfolie (Testauswertung) 176

Trennschärfe 61f.

- koeffizient 62

Type-Token-Ratio (TTR) s. Diversifikationskoeffizient

T-Werte 78, 245f.

Umgang mit Menschen, Interessen für (Test) 145

Umstrukturierungsfähigkeit 15f., 21

Umwelt/anregung 34ff., 68

- bedingungen s. Milieu-

Umzentrierung 16

Underachievement 82, 121, 132, 163

Units (Denkprodukte) 20f.

Universalität (erfaßter Intelligenz-faktoren) 132

Unterricht, leistungsdifferenzierter 81

Unterschiedsgenauigkeit (Accuracy of discrimination) 133

Urteilsfähigkeit 28

Validität 63, 72ff., (Tests) 114ff., 163f., 218

- Inhalts- 73f.
- Konstrukt- 75f.
- kriterienbezogene 74f.
- logische 73
- triviale 73
- Übereinstimmungs- 73f.
- Vorhersage- 73f.

Validitäts/koeffizient 73f.

- kriterien 73f.

Validity s. Validität

- Concurrent (Übereinstimmungs-) 73
- Construct (Konstrukt-) 73
- Content (Inhalts-) 73
- Predictive (Vorhersage-) 73

Varianz, beobachtete 64

- Fehler- 64f.
- wahre 64

V : ed 15

Vegetativ labil vs. stabil (Test) 144

Verarbeitungskapazität 28

Verbal 15f. (s. auch Fähigkeiten, verbale)

- Ability 15
- Comprehension 18, 28, 146, 214f.
- factor 136, 147, 185, 192

Verbalität s. Verbal factor

Verhalten, begabtes 12f.

- sozial-integratives 22

Verhaltensbeobachtung 83ff., 152

- Interpretation der 85f., 89ff.

Verlässlichkeit s. Reliabilität

Verhältnisskala 77

Verrechnung/Verwaltung, Interessen für (Test) 145, 154

Vertrauensbereich s. Konfidenzintervall

Visualisation 137

Visual Retention Test 140

Visuelle Denkinhalte s. figurale

Visus (Fern- vs. Nah-) 209, 212f.

- kriterien 212

Überschau (Intelligenz) 15

Übertrittsauslese 161ff.

Vocational Intelligence Scale for Adult  
Blind (VISAB) 141ff.

Vorlieben (Interessen) 145f., 154f.

Vorschule s. Bildungsförderung,  
vorschulische

Vorstellungsbesitz (Phantasie) 11

Wahrnehmungsfeld 12

– tempo 19, 137, 139

– verhalten 12

Wahrscheinlicher Fehler (WF) 68, 70

Wechsler Adult Intelligence Scale  
(WAIS) 114, 207

Wechsler Bellevue Intelligence Scale  
32, 114

Wechsler Intelligence Scale for Children  
(WISC) 119, 134, 207, 215f.

Wertende (bewertende) Denkfaktoren 19f.

Werte, ipsative 154f.

– normative 154f.

Wertpunkte (WP) 78, 118f., 245

Wesensschau 13

Wiedererkennen von Symbolen 137

Willensstärke 152

Williams Intelligence Test for Children  
with Defective Vision 100

Wissenschaftliche Interessen (Test) 145,  
145, 154

Word fluency 18, 21, 25, 136, 147

Wort/einfall 136, 139

– flüssigkeit s. Word fluency

– schatz, Grund- 215

– schatz, semantische Struktur des 40

– schatztest (WT) 115

– schatztests, Frankfurter 52, 146f.

– schatzumfang 40, 136

Zahlen/nachsprechen (ZN) 115

– reihen (ZR) 129

– Symbol-Test (ZS) 28, 116ff., 123

Zeichnen (SON) 125

Zulänglichkeit 72f.

Zuverlässigkeit s. Reliabilität

Zuwendungsbereitschaft 161

Zwei-Faktoren-Theorie 14, 17, 43, 132

Zweiseitige Fragestellung 70f.

Zwillingsforschung 13, 33f., 47

z-Wert-Skala 67, 76ff.